

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 18, 2002, 22:33:40 ; Search time 52 Seconds
(without alignments)
2901.634 Million cell updates/sec

Title: US-09-686-522C-13
Perfect score: 492
Sequence: 1 agctcgcgcgcattctcc.....agccaatttgngaagaga 492

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	57.4	11.7	7218	1	US-08-232-463-14
2	54.4	11.1	320	4	US-09-165-264-7
3	54.4	11.1	320	4	US-09-165-264-11
4	53.8	10.9	318	4	US-09-165-264-12
5	53.8	10.9	2721	6	5215881-2
6	53.8	10.9	8438	1	US-07-945-283-1
7	53.6	10.9	320	4	US-09-165-264-13
8	53.4	10.9	319	4	US-09-165-264-8
9	52.8	10.7	12001	1	US-08-458-568A-11
10	52.6	10.7	320	4	US-09-165-264-14
11	49.4	10.0	530	3	US-08-765-907A-1
12	49.4	10.0	4524	3	US-08-845-998-7
13	49.4	10.0	4524	3	US-09-206-537-7
14	49.4	10.0	4524	4	US-09-430-854-7
15	48	9.8	53526	3	US-08-658-136-2
16	48	9.8	53577	3	US-08-658-136-1
17	47.6	9.7	152331	3	US-09-128-155-16
18	47.4	9.6	2888	4	US-08-765-907A-1
19	47.2	9.6	530	3	US-08-765-907A-1
20	47.2	9.6	2823	1	US-08-398-662-4
21	47.2	9.6	2823	2	US-08-398-008A-1
22	47.2	9.6	4524	2	US-08-893-333-1
23	47.2	9.6	4524	3	US-08-845-998-7
24	47.2	9.6	4524	4	US-09-206-537-7
25	46.8	9.5	1926	4	US-09-430-854-7
26	46.8	9.5	2580	3	US-09-249-585A-2
27	46.8	9.5	2580	4	US-09-050-863-2
					US-09-359-081-2

28	46.8	9.5	5452	2	US-09-130-114-1	Sequence 1, Appli
29	46.8	9.5	9600	4	US-08-910-647-1	Sequence 1, Appli
30	46.8	9.5	9600	4	US-09-620-925-1	Sequence 1, Appli
31	46.8	9.5	10596	1	US-07-884-811-15	Sequence 15, Appli
32	46.8	9.5	10596	1	US-07-885-971-15	Sequence 15, Appli
33	46.8	9.5	10596	1	US-08-087-783A-15	Sequence 15, Appli
34	46.8	9.5	10596	1	US-08-194-088B-15	Sequence 15, Appli
35	46.8	9.5	10596	2	US-08-194-087-15	Sequence 15, Appli
36	46.8	9.5	10596	5	PCT-US93-04648-15	Sequence 15, Appli
37	46.4	9.4	1515	4	US-09-292-768-5	Sequence 5, Appli
38	45.6	9.3	289	4	US-09-007-005-17	Sequence 17, Appli
39	45.6	9.3	289	4	US-09-244-796-17	Sequence 17, Appli
40	45.6	9.3	1028	4	US-08-118-200-1	Sequence 1, Appli
41	45.6	9.3	1028	4	US-08-458-745-1	Sequence 1, Appli
42	45.6	9.3	2830	1	US-07-882-292-1	Sequence 1, Appli
43	45.6	9.3	2830	2	US-08-331-644-1	Sequence 1, Appli
44	45.6	9.3	2830	5	PCT-US93-04102-1	Sequence 1, Appli
45	45.4	9.2	2647	5	PCT-US93-06251-77	Sequence 77, Appli

ALIGNMENTS

RESULT 1
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMM
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)835-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZ9pt-Fls
; US-08-232-463-14

Query Match 11.7% Score 57.4; DB 1; Length 7218;

REGISTRATION NUMBER: 36,317

131 GCATACCTAACCCGCTCTCGCT

RY 131 GCAAAACCTAAACCCCGGCTCTCGGCTCTCGGCTCGGCGCGGCTCGCCC 181


```

RESULT 14
US-09-430-854-7/c
; Sequence 7, Application US/09430854
; Patent No. 6271019
; GENERAL INFORMATION:
; APPLICANT: Van Baren, Nicolas
; APPLICANT: Coulie, Pierre G.
; APPLICANT: De Smet, Charles
; APPLICANT: Lucas, Sophie
; APPLICANT: Boon, Thierry
; TITLE OF INVENTION: LEUKEMIA ASSOCIATED GENES
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRES:
; ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
; STREET: 600 Atlantic Avenue
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/430,854
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/845,998
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Van Amsterdam, John R.
; REGISTRATION NUMBER: 40,212
; REFERENCE/DOCKET NUMBER: L0461/7008
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)720-3500
; TELEFAX: (617)720-2441
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4524 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 174..1433
US-09-430-854-7

```

[illegible][illegible]

[illegible]

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Query	Score	Match	Length	DB	ID	Description
1	334	19.3	261	1	YHEB_CHLVI	P56160	chlorobium
2	292	16.9	271	1	SUHH_PSEAE	Q9hxi4	pseudomonas
3	287	16.6	267	1	SUHH_PASMU	Q9cnv8	pasteurella
4	279	16.1	264	1	SUHH_AQUAE	Q67791	aquifex aeol
5	272	15.7	267	1	SUHH_HAEIN	P44333	haemophilus
6	265.5	15.3	268	1	MYO3_LYCES	P44328	lycopersico
7	262	15.1	267	1	SUHH_SALTY	P58537	salmonella
8	258	14.9	267	1	SUHH_ECOLI	P22783	escherichia
9	248	14.3	267	1	SUHH_VIBCH	Q9kty5	vibrio chol
10	247.5	14.3	273	1	MYO1_LYCES	P54926	lycopersico
11	238	13.8	265	1	MYO2_LYCES	P49071	mesembryanti
12	235	13.6	270	1	MYOP_MESCR	Q49071	mesembryanti
13	234	13.5	275	1	SUHH_XYLFA	Q9pam0	xyella fast
14	229.5	13.3	256	1	SUHH_THEMEA	Q33832	thermotoga
15	228	13.2	266	1	SUHH_RHILLO	Q98f59	rhizobium l
16	221.5	12.8	265	1	SUHH_BACSU	Q45499	bacillus su
17	218	12.6	259	1	STRO_STRGR	P29785	streptomyces
18	212.5	12.3	287	1	SUHH_SYNY3	P74158	synecobact
19	211	12.2	266	1	SUHH_RHIME	Q92m71	rhizobium m
20	211	12.2	277	1	MYOP_BOVIN	P20456	bos taurus
21	209	12.1	285	1	MYOP_XENIA	P29219	xenopus lae
22	206	11.9	246	1	CSYS_ECOLI	P22255	escherichia
23	203.5	11.8	277	1	MYOP_MOUSE	O55023	mus musculus
24	202.5	11.7	261	1	SUHH_NEINB	Q9jz07	neisseria m
25	201.5	11.6	261	1	SUHH_NEIMA	Q9ju03	neisseria m
26	200	11.6	252	1	SUHH_METJA	Q57573	methanococ
27	196.5	11.4	288	1	MYO2_HUMAN	O14732	homo sapien
28	193.5	11.2	277	1	MYOP_HUMAN	P29218	homo sapien
29	193.5	11.2	277	1	MYOP_RAT	P97697	rattus norv
30	190.5	11.0	269	1	CSYS_ACTAC	P70714	actinobacil
31	190	11.0	246	1	CSYS_SALTY	P26264	salmonella
32	183.5	10.6	256	1	SUHH_CAUCR	Q9a3d5	caulobacter
33	181.5	10.5	341	1	YSNM_CAEFL	O19420	caenorhabdi


```

Db 233 ----GHSYLTSGHLVAAA-PRIVKEILNKIQ 258

RESULT 4
SUHB_AQUAE
ID SUHB_AQUAE STANDARD; PRT; 264 AA.
AC O67791;
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DE 01-MAR-2002 (Rel. 41, Last annotation update)
DE Inositol-1-monophosphatase (EC 3.1.3.25) (IMPase) (Inositol-1-
phosphatase) (I-1-Pase).
GN SUHB OR IMP2 OR AQ_1983.
OS Aquifex aeolicus.
OC Bacteria; Aquificales; Aquificaceae; Aquifex.
OX NCBI_TaxID=63363;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VF5;
X MEDLINE=98196666; PubMed=9537320;
A Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA Graham D.E., Overbeek R., Snead M.A., Keller M., Aujay M., Huber K.,
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT *The complete genome of the hyperthermophilic bacterium Aquifex
aeolicus.*;
RL Nature 392:353-358(1998).
CC -!- CATALYTIC ACTIVITY: Myo-inositol 1-monophosphate + H(2)O = myo-
CC inositol + phosphate.
CC -!- COFACTOR: Magnesium (By similarity).
CC -!- SIMILARITY: BELONGS TO THE INOSITOL MONOPHOSPHATASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE000766; AAC07753.1;
CC HSSP; P29218; LIMF.
CC InterPro: IPR000760; Inositol_P.
CC Pfam; PF00459; inositol_P; 1.
CC PROSITE; PS00629; IMP_1; 1.
CC PROSITE; PS00630; IMP_2; 1.
CC Hydrolase; Magnesium; Complete proteome.
CC KW SEQUENCE 264 AA; 29336 MW; 10F32D1D3B81B76A CRC64;

Query Match 16.1%; Score 279; DB 1; Length 264;
Best Local Similarity 28.5%; Pred. No. 1.1e-14;
Matches 75; Conservative 54; Mismatches 106; Indels 28; Gaps 7;

QY 76 ERLVEAQRADAAGEVLRKYFR--QRVEIDKEDHSPVTIADREAEAMVSVILKSFPPT 133
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
5 KYLEVAKIALAGGQVLFKFNCKVKENIEEKGDFSVYDKTSEERKEVILKFFPD 64

QY 134 HAIFGEEMWRCAENS-ADEFVWLDPIDGTFKSFITGKPLFGTLIALINGKPVIGVIDOP 192
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
65 HEVVGEMG---AEGSGSEYRWFIDPLDGTKNYINGFPFIFAVSVGLVKGEEPIVGAYLP 121

QY 193 ILRRRWIGVGKQTTLNQGEISVRSNCNLAQAYLYTTSPHLFEADA-----EDAFIR 245
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
122 YFDKLYWGAGKLGAYNGKRIKVDNESLKHAGVVGFPFSRRDISIYLNIFKDFVEY 181

QY 246 RNKVKVPLYGCDYAYALLASGFVDIVVESGLKPYDFLSLVPVIEGAGGSITDWRGDKLH 305
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
182 -GSMRRP--CAAAVDLCWVAEGIFDGMMEFEKMPWDI::ACLVILKEAGG----- 227

QY 306 WPVTAESRPTSFNNVAGDARVH 328
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
228 -VYTLVGEPFGVSDIIAGNKALH 249

```

```

RESULT 5
SUHB_HAEIN
ID SUHB_HAEIN STANDARD; PRT; 267 AA.
AC P44333;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Inositol-1-monophosphatase (EC 3.1.3.25) (IMPase) (Inositol-1-
phosphatase) (I-1-Pase).
GN SUHB OR HI0937.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RD / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kervatage A.R., Bult C.J., Tomb J.F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uterback T.R., Hanna M.C., Spriggen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT *Whole-genome random sequencing and assembly of Haemophilus
RT influenzae Rd.*;
RL Science 294:496-512(1995).
CC -!- CATALYTIC ACTIVITY: Myo-inositol 1-monophosphate + H(2)O = myo-
CC inositol + phosphate.
CC -!- COFACTOR: Magnesium (By similarity).
CC -!- SIMILARITY: BELONGS TO THE INOSITOL MONOPHOSPHATASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U32775; AAC22545.1;
CC HSSP; P29218; 2HHM.
CC TIGR; HI0937;
CC InterPro: IPR000760; Inositol_P.
CC Pfam; PF00459; inositol_P; 1.
CC PRINTS; PR00378; INOSPHPTASE.
CC PROSITE; PS00629; IMP_1; 1.
CC PROSITE; PS00630; IMP_2; 1.
CC Hydrolase; Magnesium; Complete proteome.
CC KW SEQUENCE 267 AA; 29499 MW; DF974E0DAB0CCEA6 CRC64;

Query Match 15.7%; Score 272; DB 1; Length 267;
Best Local Similarity 30.8%; Pred. No. 3.9e-14;
Matches 84; Conservative 44; Mismatches 113; Indels 32; Gaps 8;

QY 78 LVEAQRADAAGEVLRKYFR--VEIIDKEDHSPVTIADREAEAMVSVILKSFPPT 135
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
4 MLNIAIRARKAGNVIANKYERRDAIESTQKGINDYVTNVDRKASEAETIEVIRKSPDHT 63

QY 136 IFGEENGWRCAENSADFVWVLDPIDGTFKSFITGKPLFGTLIALHNGKPVIGVIDOPILR 195
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
64 IITEETG-AIEGKSDSDQVWIDPLDGTNFMVGLPHFSVSIARVKNRTEGVGVYDPTRN 122

QY 196 ERWIGVGKQTTLNQGEISVRSNCNLAQAYLYTTSPLH-----LFEADAEDAF-I 243
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
123 ELFTAVRGEAGAKLNEVRUKRDKREIQSGSILATGFPFKOPKLMPAQAFAMNNAIEDAADF 182

QY 244 RVRNKVKVPLYGCDYAYALLASGFVDIVVESGLKPYDFLSLVPVIEGAGGSITDWRGDK 303

```

```

Db 183 RRTGSAALD-----CY-----VASNRIDGYFEMGLKAWDCAAGDLIVREAGGLVCFD--- 231
QY 304 LHPVTAESRPTSNVVAAGDARVHKALDALR 336
Db 232 -----AGNSYLRSNGIIIAA-PSRVIREKMLNKR 258

RESULT 6
MYO3_LYCES
ID MYO3_LYCES STANDARD; PRT; 268 AA.
AC P54928;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myo-inositol-1(or 4)-monophosphatase 3 (EC 3.1.3.25) (IMPase 3) (IMP
GN IMP3;
OS Lycopersicon esculentum (Tomato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
NCBI_TaxID=4081;
[1]
SEQUENCE FROM N.A.
STRAIN=CV_VENT CHERRY;
MEDLINE=96351935; PubMed=8718627;
RA Gillaspie G.E., Keddle J.S., Oda K., Gruissem W.;
RT "Plant inositol monophosphatase is a lithium-sensitive enzyme encoded
RL by a multigene family.";
PL Plant Cell 7:2175-2185(1995).
CC -!- FUNCTION: IT IS RESPONSIBLE FOR THE PROVISION OF INOSITOL REQUIRED
CC FOR SYNTHESIS OF PHOSPHATIDYLINOSITOL AND POLYPHOSPHOINOSITIDES.
CC -!- CATALYTIC ACTIVITY: Myo-inositol 1-monophosphate + H(2)O = myo-
CC inositol + phosphate.
CC -!- COFACTOR: REQUIRES MAGNESIUM (BY SIMILARI Y).
CC -!- ENZYME REGULATION: INHIBITED BY LI(+).
CC -!- PATHWAY: KEY ENZYME OF THE PHOSPHATIDYL INOSITOL SIGNALING
CC PATHWAY.
CC -!- SIMILARITY: BELONGS TO THE INOSITOL MONOPHOSPHATASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U39059; AAB19031.1; -
CC HSP; P29218; 2HUM.
CC InterPro; IPR000760; Inositol_P.
CC Pfam; PF00459; inositol_P; 1.
CC PRINTS; PR00378; INOSPHPTASE.
CC PROSITE; PS00629; IMP_1; 1.
CC PROSITE; PS00630; IMP_2; 1.
CC Hydroxylase; Magnesium; Lithium; Multigene family.
CC SEQUENCE 268 AA; 29054 MW; BCB635A029804901 CRC64;

Query Match 15.3%; Score 265.5; DB 1; Length 268;
Best Local Similarity 30.2%; Pred. No. 1.2e-13;
Matches 83; Conservative 48; Mismatches 117; Indels 27; Gaps 8;

QY 74 ATERLVEAORAAAGAEVLRKYFRQRIEIIIDKEDHSPTVIADREAEAMVSVILKSSEPT 133
Db 6 SVEQFLDAVEAAKAGEIIRGFFKTHVEIKGMVDLVTETDKACEDFIHNLKQRFPS 65

QY 134 HAIFGEWNRCA--ENSADFVWVLDPIIDGTSKFSITGKPLFTGLTALLHNGKPVICVTDQ 191
Db 66 HKFIEETTAACCNFEELDEPTWIVDLDGTTNFVHGFPFVCSVIGLTIKKPTGVGVYN 125

QY 192 PILRERWICVDCQKOTTLNGQEIISVRSCNLLAQAYLYT---TSPHLFEADA-----EDAFI 243

```

```

Db 126 PIIDELFTGIDGAFNGPKPIKVSQSESLVKALLATEAGTNRDKLVVDATTGRINSLF 185
QY 244 RVRNKVKVPLYGDCYAYAL--LASGFVDIVVESGL-KPYDFLSLVPIVIEGAGSGITDWR 300
Db 186 KVRB-----LRMGSCSALNLCGVACGKLDLFYELEFGPMDVAGGAVIVKEAGGFVDP 240
QY 301 GDKLHPVTAESRPTSNVVAAGDARVHKALDAL 335
Db 241 GSE--FDLTARR-----VAATNAHLKDAFIKAL 266

RESULT 7
SUHB_SALTY
ID SUHB_SALTY STANDARD; PRT; 267 AA.
AC P58537;
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Inositol-1-monophosphatase (EC 3.1.3.25) (IMPase) (Inositol-1-
GN phosphatase) (I-1-Pase).
GN SUHB OR STM2546.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
NCBI_TaxID=602;
[1]
SEQUENCE FROM N.A.
STRAIN=LT2 / SGSC1412 / ATCC 700720;
MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea B., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2.";
RL Nature 413:852-856(2001).
CC -!- CATALYTIC ACTIVITY: Myo-inositol 1-monophosphate + H(2)O = myo-
CC inositol + phosphate.
CC -!- COFACTOR: Magnesium (By similarity).
CC -!- SIMILARITY: BELONGS TO THE INOSITOL MONOPHOSPHATASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE008815; AAL21440.1; -
CC StyGene; SG?????; suhb.
CC PROSITE; PS00629; IMP_1; 1.
CC PROSITE; PS00630; IMP_2; 1.
CC Hydroxylase; Magnesium; Complete proteome.
CC SEQUENCE 267 AA; 29158 MW; FC4608A8298BB334 CRC64;

Query Match 15.1%; Score 262; DB 1; Length 267;
Best Local Similarity 30.0%; Pred. No. 2.3e-13;
Matches 80; Conservative 44; Mismatches 123; Indels 20; Gaps 7;

QY 78 LVEAORAAAGAEVLRKYFR--QRVEIIDKEDHSPTVIADREAEAMVSVILKSFFTHA 135
Db 4 MLTIARAARKAGNVIKAKNYETPDPAVESQKNSDFVTNVDKAAEAVIIDTIRKSYPOHT 63

QY 136 IFGEWNRCAENSADFVWVLDPIIDGTSKFSITGKPLFTGLTALLHNGKPVIGVDOPTLR 195
Db 64 ITESGHEVCTDQ-DVQWVIDPLDGTTFNFKRPLHFAVIAVRKRGTEAVVYDPMRN 122

QY 196 ERWICVDCQKOTTLNGQEIISVRSCNLLAQAYLYT--TSPHLFEA--DAEDAFIRNRKVKVPL- 253

```

```

CC -!- CATALYTIC ACTIVITY: Myo-inositol 1-monophosphate + H(2)O = myo-
CC inositol + phosphate.
CC -!- COFACTOR: Magnesium.
CC -!- SUBUNIT: Monomer.
CC -!- SIMILARITY: BELONGS TO THE INOSITOL MONOPHOSPHATASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```


DR PROSITE: PS00630; IMP_2: 1;
KW Hydrolase; Magnesium; Lithium; Multigene family.
SQ SEQUENCE 273 AA; 29558 MW; 92DD532E359E0FD CRC64;

	Query Match	14.3%;	Score 247.5;	DB 1;	Length 273;
	Best Local Similarity	28.9%;	Pred. NO. 3.2e-12;		
Matches	80;	Conservative	53;	Mismatches 111;	Indels 33; Gaps 10;
QY	76	ERLVEVAQRADAAGEVLRKYFRQVRVEIADKEDHSPPVTIADRAEEAEAMVSILKSPPTHA	135		
Db		: : : : :			
	8	EELGVAVDAAKRAGEIRKKGFHETKIUVHKCKQDVLVTEDTKACEDLI FNHLQAHPESHK	67		
QY	136	IFGEENWCRAENSADH----VVWLDPIDINGTKSFITGPKPLFCTLIALLHNKPKWIGVI	189		
Db		:			
	68	FIGEET-----SAATGDFDLTLEPTWIVDPDPTNFVHGFPSCVSGITLGIPATVGW	123		
QY	190	DQPILRRWIGVDGKQTTLNQCQISVRSCMLLAQAYL----YTTSPHLFEADAEDAFIRV	245		
Db		: : : : : : :			
	124	YDPIIDELEFTINGKGAYLNKGP IKVYSQSSELVKSLUGTEVGTTDRNL---TVETTTRRI	180		
QY	246	RN--KVKVPPLYGDCDYAYAL--LASGFVDIVVESGL-RPYDFLSLPVFIEGAGGSITDW	299		
		: : : : : : :			

```

QY 300 RGDKLHWPTAESRPTSFNVVAAGDARVHKALDALR 336
      | : : | : : | : : | : : | : : | : : | : : | : : |
Db 240 SGSE--FDITSQR-----VAATNPHLKEAFVEALQ 267

RESULT 11
MYO2.LYCES
ID MYO2.LYCES STANDARD; PRT; 265 AA.
AC P54327;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myo-inositol-(or 4)-monophosphatase 2 (EC 3.1.3.25) (IMP
DE 2) (Enzitol monophosphatase)

```


DE Inositol-1-monophosphatase (EC 3.1.3.25) (IMPase) (Inositol-1-phosphatase) (I-1-Pase).

GN SUHB OR XF2476

OS Xylella fastidiosa.

OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;

OC Xylella.

OX NCBI_TaxID=2371;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=9a5C;

RX MEDLINE=2055717; PubMed=10910347;

RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,

RA Alvares R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,

RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,

RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carver H.,

RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,

RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorfi H.,

RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,

RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furian L.R.,

RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,

RA Ho P.L., Hoheisel J.D., Junqueira M.L., Kempf E.L., Kitajima J.,

RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,

RA Lemos E.G.M., Lenos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,

RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,

RA Marques M.V., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,

RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,

RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.P.S.,

RA Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,

RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,

RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,

RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,

RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,

RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,

RA da Silva J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,

RA de Souza A.P., Terenzi M.F., Truffi D., Tsa S.M., Tshako M.H.,

RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,

RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.

RT "The genome sequence of the plant pathogen Xylella fastidiosa.";

RL Nature 406:151-159(2000).

CC -!- CATALYTIC ACTIVITY: Myo-inositol 1-monophosphate + H(2)O = myo-inositol + phosphate.

CC -!- COFACTOR: Magnesium (By similarity).

CC -!- SIMILARITY: BELONGS TO THE INOSITOL MONOPHOSPHATASE FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

CC EMBL; AE004055; AAF85274.1; .

DR HSSP; P29218; 11MB.

DR InterPro; IPR000760; Inositol_P.

DR Pfam; PF00459; inositol_P; 1.

DR PROSITE; PS00629; IMP_1; 1.

DR PROSITE; PS00630; IMP_2; 1.

KW Hydrolase; Magnesium; Complete proteome.

SQ SEQUENCE 275 AA; 30233 MW; 74F622ACC4E082C CRC64;

Query Match 13.5% Score 234; DB 1; Length 275;

Best Local Similarity 28.4%; Pred. No. 3 6e-11;

Matches 73; Conservative 51; Mismatches 117; Indels 16; Gaps 8;

Oy 79 VEVAQRADAGEVLRYKFR--QRVEIDKEDHSPTVIADREAEAMVSVILKSFTTHAI 136

Db 6 VNIWKAARSAGNYLLRHINKLETLHVIOKSRMDYASDVDEMAEKVVKELKRAYPEYGI 65

Oy 137 FGEIENGWRCASNDYFVWLDIPIDGTSFITGKPLFCTLLALHNCKPVGIVGDPIRE 196

Db 66 LGEEGG---LOGNIRIMWVTDPLDCTSNYLRGPPHYICISIALVENGEPTDAVIFDPLRNE 122

Oy 197 RWIGVDGKQTTLNQGEISVRSNLLAQAYLYT-TSP-HLFEADAE-----DAFIRVRNKVK 250

Db 123 LFTASRGAGAILNKRKIRVANRKDLNGTMTLNTGSPRSRAHAQLKCYDALLMQAEDIR 182

Oy 251 VPLYGCDYAYALLASGFVDLVESGLKPKYDFLSLVPIVIEGAGSITDWRGDKLHPVTA 310

Db 183 --RGSAAILDYVACGRADYFEAGIKVMDVAAGMLLVREAGGYGVDFKGADA--PRMD 238

Oy 311 ESRPTSFNVVAAGDARV 327

Db 239 DKGPESCQLV-AGNIKV 254

RESULT 14

SUHB_THEME

ID SUHB_THEME STANDARD; PRT; 256 AA.

AC O33832;

DT 15-DEC-1998 (Rel. 37, Created)

DT 15-DEC-1998 (Rel. 37, Last sequence update)

DT 01-MAR-2002 (Rel. 41, Last annotation update)

DE Inositol-1-monophosphatase (EC 3.1.3.25) (IMPase) (Inositol-1-phosphatase) (I-1-Pase).

DE SUHB OR TM1415.

GN Thermotoga maritima.

OS Bacteria; Thermotogales; Thermotoga.

OX NCBI_TaxID=2336;

RP [1]

RP SEQUENCE FROM N.A.

RC STRAIN=MSB8 / DSM 3109;

RC MEDLINE=98386670; PubMed=9720201;

RA Liebl W., Brem D., Gotschlich A.;

RT "Analysis of the gene for beta-fructosidase (invertase, inulinase) of the hyperthermophilic bacterium Thermotoga maritima, and characterisation of the enzyme expressed in Escherichia coli.";

RL Appl. Microbiol. Biotechnol. 50:55-64(1998).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=MSB8 / DSM 3109;

RC MEDLINE=99287316; PubMed=10360571;

RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,

RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,

RA McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,

RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,

RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,

RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;

RT "Evidence for lateral gene transfer between Archaea and Bacteria from genome sequence of Thermotoga maritima.";

RL Nature 399:323-329(1999).

RN [3]

RP CHARACTERIZATION.

RX MEDLINE=99437848; PubMed=10508089;

RA Chen L., Roberts M.F.;

RT "Characterization of a tetrameric inositol monophosphatase from the hyperthermophilic bacterium Thermotoga maritima.";

RL Appl. Environ. Microbiol. 65:4559-4567(1999).

CC -!- FUNCTION: Displays a 20-fold higher rate of hydrolysis of the D isoform of inositol 1-phosphate than of the L isoform.

CC -!- CATALYTIC ACTIVITY: Myo-inositol 1-monophosphate + H(2)O = myo-inositol + phosphate.

CC -!- COFACTOR: Magnesium.

CC -!- SUBUNIT: Homotetramer.

CC -!- SIMILARITY: BELONGS TO THE INOSITOL MONOPHOSPHATASE FAMILY.

CC -!- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A FRAMESHIFT IN POSITION 22.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 18, 2002, 22:58:06 : Search time 52 Seconds
(without alignments)
3496.153 Million cell updates/sec

Title: US-09-686-522c-13
Perfect score: 492
Sequence: 1 agctcgcagcgtcattctcc.....agccaatttgngagaga 492

Scoring table: IDENTITY_NUC
Gap 10.0 , Gapext 1.0

Searched: 328367 seqs, 184756068 residues

otal number of hits satisfying chosen parameters: 656734

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : Published Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	56.4	11.5	9953	10	US-09-764-847-1336
2	56.4	11.5	11189	10	US-09-764-847-1335
3	55.2	11.2	367	10	US-09-878-574-2578
4	54.8	11.1	987	10	US-09-804-682-29
5	53.8	10.9	1064	10	US-09-804-682-29
6	52.4	10.7	437	10	US-09-960-352-12835
7	51.8	10.5	1614	12	US-10-023-529-45
8	51.8	10.5	1614	12	US-10-023-523-45
9	51.8	10.5	2283	9	US-09-938-842A-1691
10	51.8	10.5	2283	10	US-09-887-576-645
11	51.8	10.5	12425	12	US-10-023-529-50
12	51.8	10.5	12425	12	US-10-023-523-50
13	51.2	10.4	5025	10	US-09-960-253-176
14	49.8	10.1	277	10	US-09-878-574-14649
15	49.8	10.1	2561	12	US-10-023-529-48
16	49.8	10.1	2561	12	US-10-023-523-48
17	48.4	9.8	2481	10	US-09-894-998-35
18	48.2	9.8	326	10	US-09-864-761-17105
19	48.2	9.8	474	10	US-09-864-761-11284

c 20	48.2	9.8	477	10	US-09-864-761-286	Sequence 286, App
c 21	48.2	9.8	477	10	US-09-864-761-287	Sequence 287, App
c 22	47.6	9.7	463	10	US-09-864-761-5509	Sequence 5509, App
c 23	47.6	9.7	152331	9	US-10-095-407-16	Sequence 16, Appl
c 24	47.4	9.6	1065	10	US-09-804-682-33	Sequence 33, Appl
c 25	47.2	9.6	936	10	US-09-974-300-836	Sequence 836, App
c 26	47.2	9.6	42999	9	US-09-799-462A-17	Sequence 17, Appl
c 27	47.2	9.6	42999	9	US-10-125-767-17	Sequence 17, Appl
c 28	46.8	9.5	410	10	US-09-983-965-4888	Sequence 4888, App
c 29	46.8	9.5	267	10	US-09-864-761-27984	Sequence 27984, App
c 30	46.6	9.5	3809	12	US-10-001-870-68	Sequence 68, Appl
c 31	46.4	9.4	4450	10	US-09-764-887-497	Sequence 497, App
c 32	45.8	9.3	404	10	US-09-983-965-5091	Sequence 5091, App
c 33	45.8	9.3	57130	10	US-09-835-081-3	Sequence 3, Appl
c 34	45.2	9.2	659158	9	US-09-771-208-20	Sequence 20, Appl
c 35	44.8	9.1	501	10	US-09-878-574-4421	Sequence 4421, App
c 36	44.8	9.1	531	10	US-09-864-761-22281	Sequence 22281, App
c 37	44.8	9.1	1602	10	US-09-833-790-417	Sequence 417, App
c 38	44.8	9.1	4454	10	US-09-764-887-496	Sequence 496, App
c 39	44.8	9.1	174424	10	US-09-967-768A-314	Sequence 314, App
c 40	44.6	9.1	1203	9	US-09-879-312-1	Sequence 1, Appl
c 41	44.4	9.0	367	10	US-09-878-574-2578	Sequence 2578, App
c 42	44.4	9.0	3901	10	US-09-834-975-1043	Sequence 1043, App
c 43	44.2	9.0	106	10	US-09-960-352-5165	Sequence 5165, App
c 44	44.2	9.0	152331	9	US-10-095-407-16	Sequence 16, Appl
c 45	43.8	8.9	266	10	US-09-960-352-3882	Sequence 3882, App

ALIGNMENTS

RESULT 1
US-09-764-847-1336
; Sequence 1336, Application US/09764847
; Patent No. US20020132767A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC009
; CURRENT APPLICATION NUMBER: US/09/764.847
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 2003
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1336
; LENGTH: 9953
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-847-1336

Query Match 11.5%; Score 56.4; DB 10; Length 9953;
Best Local Similarity 58.2%; Pred. No. 0.0002;
Matches 99; Conservative 0; Mismatches 71; Indels 0; Gaps 0;
Qy 21 GCCCGCATTTAAACCTTGTCCGACACGCGCGGAGTCTCCCAATGCTCTCT 80
Db 3991 GCTTTGAGCTGGAGGCTTCTGCGCGCAGCGCGGCTCTCTCTCTCTCTCT 4050
Qy 81 CTTCCTCTCCACCATCTGGCCACCTGCGCGGCTCTCCCGGCTCTCCGCAACCTA 140
Db 4051 CTTCT 4110
Qy 141 ACCCGGCTCTGCGCTCTCTGCGCTCTGCGCGGCTCTGCGCGGCTCTGCTTC 190
Db 4111 CCT 4160

RESULT 2
US-09-764-847-1335
; Sequence 1335, Application US/09764847
; Patent No. US20020132767A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.


```
RESULT 5
US-09-804-682-29
; Sequence 29, Application US/09804682
; Patent No. US20020106765A1
; GENERAL INFORMATION:
; APPLICANT: Kinders, Robert J.
; TITLE OF INVENTION: PAL-18 POLYPEPTIDES, NUCLEIC ACIDS
; TITLE OF INVENTION: ENCODING THE SAME AND METHODS FOR SCREENING FOR OR
; TITLE OF INVENTION: MODULATING THE SAME
; FILE REFERENCE: 130001.406
; CURRENT APPLICATION NUMBER: US/09/804.682
; CURRENT FILING DATE: 2001-03-12
; NUMBER OF SEQ ID NOS: 174
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 29
; LENGTH: 1064
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 5, 6, 16, 21, 24, 25, 33, 39, 72, 110, 209, 214, 231, 232,
; LOCATION: 235, 237, 238, 244, 245, 246, 25, 282, 292, 297, 306, 319,
; LOCATION: 321, 323, 330, 340, 349, 354, 355, 363, 372, 376, 378,
; LOCATION: 397, 405, 432, 437, 454, 455, 457, 458, 459, 468, 470
; OTHER INFORMATION: n = A,T,C or G
; NAME/KEY: misc_feature
; LOCATION: 485, 487, 488, 494, 496, 499, 511, 524, 527, 552, 557, 562,
; LOCATION: 583, 600, 611, 613, 623, 624, 652, 654, 674, 681, 687, 691,
; LOCATION: 694, 701, 713, 716, 720, 721, 725, 731, 734, 735, 739, 743,
; LOCATION: 744, 781, 782, 785, 789, 799, 803, 821, 823, 847, 852
; OTHER INFORMATION: n = A,T,C or G
; NAME/KEY: misc_feature
; LOCATION: 858, 878, 884, 886, 896, 897, 901, 917, 926, 932, 939, 948,
; LOCATION: 957, 961, 963, 981, 991, 993, 1001, 1002, 1005, 1011, 1018,
; LOCATION: 1043, 1047, 1049, 1051, 1054, 1056
; OTHER INFORMATION: n = A,T,C or G
US-09-804-682-29
Query Match
Best Local Similarity 10.9%; Score 53.8; DB 10; Length 1064;
Matches 91; Conservative 0; Mismatches 88; Indels 0; Gaps 0;
QY 3 CTCGAGCGTCAATCTCCGCGCCGACATTTAAACCTTGCTCCGACAAACGCGCGCGACT 62
DB 677 CTCGCGCCGCGCTGCTCCGCGCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 736
QY 63 CTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 122
DB 737 CCNCCNCCNCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 796
QY 123 TCGCGCTCCGCAAAACCTTAACCGCGCTGCTCCGCGCGCGCGCGCGCGCGCGCGCG 181
DB 797 CCNCCNCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 855
RESULT 6
US-09-960-352-12835
; Sequence 12835, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
US-09-960-352-12835
; SEQ ID NO 12835
; LENGTH: 437
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 55-LIB2809-007-Q1-E1-F4
US-09-960-352-12835
Query Match
Best Local Similarity 10.7%; Score 52.4; DB 10; Length 437;
Matches 101; Conservative 0; Mismatches 81; Indels 0; Gaps 0;
QY 16 TCTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 75
DB 227 TCTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 286
QY 76 CTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 135
DB 287 CCCTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 346
QY 136 CCCTAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 195
DB 347 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 406
QY 196 CT 197
DB 407 CT 408
RESULT 7
US-10-023-529-45
; Sequence 45, Application US/10023529
; Patent No. US20020129388A1
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; TITLE OF INVENTION: ATHEROSCLEROSIS
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/10/023,529
; CURRENT FILING DATE: 2001-12-17
; PRIOR APPLICATION NUMBER: 09/616,289
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/517,849
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 60/031,930
; PRIOR FILING DATE: 1996-11-27
; PRIOR APPLICATION NUMBER: US 60/048,547
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 45
; LENGTH: 1614
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1614)
US-10-023-529-45
Query Match
Best Local Similarity 10.5%; Score 51.8; DB 12; Length 1614;
Matches 189; Conservative 0; Mismatches 193; Indels 2; Gaps 2;
QY 3 CTCGAGCGTCAATCTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 62
DB 404 CACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 463
QY 63 CCTCCCCAATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 122
```

```
RESULT 5
US-09-804-682-29
; Sequence 29, Application US/09804682
; Patent No. US20020106765A1
; GENERAL INFORMATION:
; APPLICANT: Kinders, Robert J.
; TITLE OF INVENTION: PAL-18 POLYPEPTIDES, NUCLEIC ACIDS
; TITLE OF INVENTION: ENCODING THE SAME AND METHODS FOR SCREENING FOR OR
; TITLE OF INVENTION: MODULATING THE SAME
; FILE REFERENCE: 130001.406
; CURRENT APPLICATION NUMBER: US/09/804.682
; CURRENT FILING DATE: 2001-03-12
; NUMBER OF SEQ ID NOS: 174
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 29
; LENGTH: 1064
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 5, 6, 16, 21, 24, 25, 33, 39, 72, 110, 209, 214, 231, 232,
; LOCATION: 235, 237, 238, 244, 245, 246, 25, 282, 292, 297, 306, 319,
; LOCATION: 321, 323, 330, 340, 349, 354, 355, 363, 372, 376, 378,
; LOCATION: 397, 405, 432, 437, 454, 455, 457, 458, 459, 468, 470
; OTHER INFORMATION: n = A,T,C or G
; NAME/KEY: misc_feature
; LOCATION: 485, 487, 488, 494, 496, 499, 511, 524, 527, 552, 557, 562,
; LOCATION: 583, 600, 611, 613, 623, 624, 652, 654, 674, 681, 687, 691,
; LOCATION: 694, 701, 713, 716, 720, 721, 725, 731, 734, 735, 739, 743,
; LOCATION: 744, 781, 782, 785, 789, 799, 803, 821, 823, 847, 852
; OTHER INFORMATION: n = A,T,C or G
; NAME/KEY: misc_feature
; LOCATION: 858, 878, 884, 886, 896, 897, 901, 917, 926, 932, 939, 948,
; LOCATION: 957, 961, 963, 981, 991, 993, 1001, 1002, 1005, 1011, 1018,
; LOCATION: 1043, 1047, 1049, 1051, 1054, 1056
; OTHER INFORMATION: n = A,T,C or G
US-09-804-682-29
Query Match
Best Local Similarity 10.9%; Score 53.8; DB 10; Length 1064;
Matches 91; Conservative 0; Mismatches 88; Indels 0; Gaps 0;
QY 3 CTCGAGCGTCAATCTCCGCGCCGACATTTAAACCTTGCTCCGACAAACGCGCGCGACT 62
DB 677 CTCGCGCCGCGCTGCTCCGCGCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 736
QY 63 CTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 122
DB 737 CCNCCNCCNCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 796
QY 123 TCGCGCTCCGCAAAACCTTAACCGCGCTGCTCCGCGCGCGCGCGCGCGCGCGCGCG 181
DB 797 CCNCCNCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 855
RESULT 6
US-09-960-352-12835
; Sequence 12835, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
```

```

Db 464 CCGCCACAGCGCCCTCGCTTGGCCCGCGCAGCGGGC-CCCCGCGGACGCGGGC 522
QY 123 TCGCCTCCGCAAAACCTTAACCCCGCTCTCGCTTCCGCTTCCGCGCGCTTCCCGC 182
Db 523 CGCGCCCTTGGCCGCGCGCGCGCGCGCGCGCTTCCCGCGCGCTTGGCGCGCGCGC 582
QY 183 TGTGCTCGCGGCTTTAGCGCGAGTGGCGCCGAGTGGCGCGGATGAGTACGGTTAGGGCCTCT 242
Db 583 GGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 642
QY 243 TCGCGCTGGGGCGCGCGCGCGGAGCTCGCGAGTGGGGAGTTGGCGAC-GGAGCGG 301
Db 643 CCACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 702
QY 302 CTGGTGGAGTGGCGCAACGGGGCGGAGCGCTGCTGGGGAGTGTCTCANGAAGTACTTC 361
Db 703 CCGGAGGGGGCGCGGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 762
QY 362 CGCGAGCGGTTGAGATCATCGAC 385
Db 763 GTGCGTACCTCGGGGCGACGCG 786

RESULT 8
US-10-023-523-45
; Sequence 45, Application US/10023523
; Patent No. US20020152485A1
; GENERAL INFORMATION:
; APPLICANT: Lees, Ann M.
; APPLICANT: Lees, Robert S.
; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; FILE OF INVENTION: ATHEROSCLEROSIS
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/10/023,523
; CURRENT FILING DATE: 2001-12-17
; PRIOR APPLICATION NUMBER: US/09/616,289
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/517,849
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 60/031,930
; PRIOR FILING DATE: 1996-11-27
; PRIOR APPLICATION NUMBER: US 60/048,547
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 45
; LENGTH: 1614
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1614)
US-10-023-523-45

Query Match 10.5%; Score 51.8; DB 12; Length 1614;
Best Local Similarity 49.2%; Pred. No. 0.0014;
Matches 189; Conservative 0; Mismatches 193; Indels 2; Gaps 2;

QY 3 CTCGAGCGCTCATTTCCGCGCGGACATTTAAACCTTGTCTCCGACACACCGCGCGCGACT 62
Db 404 CACCGCGCGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 463
QY 63 CCGTCCCAATGCTCTCTCTTCTCTCTTCCACCCACTCGCGCCACCTCGCGCTTCCCGCGC 122
Db 464 CCGCCACAGCGCGCGCGCTGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 522
QY 123 TCGCCTCCGCAAAACCTTAACCCCGCTCTCGCTTCCGCTTCCGCGCGCGCTTCCCGCGC 182
```

```

Db 523 GCGCCCTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTTCCCGCGCGTGGCGCGCGCC 582
QY 183 TGTGCTTCCGCGCTTGTAGCGCGAGTGGCGCGAGTGGCGCGGATGAGTACGGTTAGGGCCTCT 242
Db 583 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 642
QY 243 TCGCGCTGGGGCGCGCGCGGAGTGGCGCGGAGTGGCGAGTGGGGAGTTGGCGAC-GGAGCGG 301
Db 643 CCACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 702
QY 302 CTGGTGGAGTGGCGCAACGGGGCGGAGCGCTGCTGGGGAGTGTCTCANGAAGTACTTC 361
Db 703 CCGGAGGGGGCGCGGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 762
QY 362 CCGCAGCGGTTGAGATCATCGAC 385
Db 763 GTGCGTACCTCGGGGCGACGCG 786

RESULT 9
US-09-938-842A-1691
; Sequence 1691, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINI
; FILE OF INVENTION: SAME, AND METHODS OF USE
; FILE REFERENCE: SCRIPI300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 1691
; LENGTH: 2283
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-1691

Query Match 10.5%; Score 51.8; DB 9; Length 2283;
Best Local Similarity 51.5%; Pred. No. 0.0016;
Matches 119; Conservative 0; Mismatches 112; Indels 0; Gaps 0;

QY 22 CCGCGACATTTAAACCTTGTCTCCCGACAAACCGCGCGACTCTCCCAATGCTCTCCTC 81
Db 1295 CTCGCGCTGTTATTCTCCCGCTCTCTCCACCGCGCACACTCTCCCGGTATATCTCCTC 1354
QY 82 TTCCTCTCCACCGACTCGCGACCTCGCCCTTCCCGCGCGCTCGCGCTCCGCAAAACCTTAA 141
Db 1355 CACCACCACCGCGCGCGCGCGCTCTCTCCCGCAGTATATCTCTCCACCAACACACCGC 1414
QY 142 CCGCGCTCTCGCTCTCTCCCGCTCCCGCGCGCGCGCTCGCGCGCGCTCGCTCGCGGTCTT 201
Db 1415 CCGCAGCGCTCTCTCGCGCGCGCTCTCTCTCCCGCGCGCGCTCTCTCTCCCGCGCTC 252
QY 202 CCGGAGTGGCGCGCGCGCGGATGAGTACGGTTAGGGCGCTTGGCGCGCGCTG 1474
Db 1475 CGCCAGTCTACTCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1525

RESULT 10
US-09-887-576-645
; Sequence 645, Application US/09887576
; Patent No. US2002014047A1
; GENERAL INFORMATION:
```


RESULT 14

US-09-878-574-14649

Sequence 14649, Application US/09878574

Patent No. US20020110548A1

GENERAL INFORMATION:

APPLICANT: Byrum, Joseph R.

APPLICANT: La Rosa, Thomas J.

APPLICANT: Thompson, Michael D.

TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with

TITLE OF INVENTION: Plants

FILE REFERENCE: 38-21(15401)B

CURRENT APPLICATION NUMBER: US/09/878,574

CURRENT FILING DATE: 2001-12-21

PRIOR APPLICATION NUMBER: 09/333,535

PRIOR FILING DATE: 1999-06-14

NUMBER OF SEQ ID NOS: 15775

SEQ ID NO 14649

LENGTH: 277

TYPE: DNA

ORGANISM: Glycine max

OTHER INFORMATION: Clone ID: 701069068H1

US-09-878-574-14649

Query Match 10.1%; Score 49.8; DB 10; Length 277;

Best Local Similarity 59.1%; Pred. No. 0.0026;

Matches 101; Conservative 0; Mismatches 69; Indels 1; Gaps 1;

QY 297 AGCGGCTGGTGGAGGTGGCGCAACGGCGGGGAGCGTCTGGGGAGGTGCTCANGAAGT 356

Db 106 ATCGCTTCGGCGAGGTTCGGTTACAAAGTCGCGGATGCTGGCGGAGAGTATTCCGCAAT 165

QY 357 ACTTCCGCCAGCGGGTTGAGATCATCGACAAAGAGGACACAGTCTCTTACAATTGCAG 416

Db 166 ACTTCAGAAAAAATCTCGACGTTATTACAAACAATGATCTCAGTCCAGTACCAATTGCAG 225

QY 417 ATAGAAGAAGCAGAGAGCAATGNGTCACTTATCTAGAGAGCTTCCT 467

Db 226 CT-CAATCTGTGAGGAGGTATGTTTCAATCATACTAGCAATTCCTT 275

RESULT 15

US-10-023-529-48

Sequence 48, Application US/10023529

Patent No. US20020129388A1

GENERAL INFORMATION:

APPLICANT: Lees, Ann M.

APPLICANT: Lees, Robert S.

APPLICANT: Law, Simon W.

APPLICANT: Arjona, Anibal A.

TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING

TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING

TITLE OF INVENTION: ATHEROSCLEROSIS

FILE REFERENCE: 10797-004001

CURRENT APPLICATION NUMBER: US/10/023,529

CURRENT FILING DATE: 2001-12-17

PRIOR APPLICATION NUMBER: 09/616,289

PRIOR FILING DATE: 2000-07-14

PRIOR APPLICATION NUMBER: US 09/517,849

PRIOR FILING DATE: 2000-03-02

PRIOR APPLICATION NUMBER: US 08/979,608

PRIOR FILING DATE: 1997-11-26

PRIOR APPLICATION NUMBER: US 60/031,930

PRIOR FILING DATE: 1996-11-27

PRIOR APPLICATION NUMBER: US 60/048,547

PRIOR FILING DATE: 1997-06-03

NUMBER OF SEQ ID NOS: 53

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 48

LENGTH: 2561

TYPE: DNA

ORGANISM: Oryctolagus cuniculus

FEATURE:

GenCore version 5.1.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 18, 2002, 19:29:19 ; Search time 2102 Seconds
(without alignments)
3790.761 Million cell updates/sec

Title: US-09-686-522c-13

Perfect score: 492

Sequence: 1 agctccgagcgtcattctcc.....agccaatttggngaagaga 492

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

otal number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba:*
2: em_estin:*
3: em_estin:*
4: em_estin:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	297	60.4	505	12	BG103856
2	297	60.4	533	12	BG053327
3	177.2	36.0	444	14	BQ240879
4	176	35.8	588	13	BI778441
5	175	35.6	863	13	BI951067
6	171.2	34.8	598	13	BM370373

7	168.4	34.2	483	9	AU093884
8	168.4	34.2	864	10	BE418829
9	161	32.7	766	12	BF631396
10	155.2	31.5	443	14	BQ467759
11	148.8	30.2	404	10	BE498543
12	145	29.5	667	13	BI959675
13	141.8	28.8	631	12	BG414965
14	119.2	24.2	759	14	BQ801093
15	115.2	23.4	745	14	BQ608457
16	90	18.3	528	10	BE443955
c	83.6	17.0	974	17	BI157536
18	79.4	16.1	974	17	BI157536
19	78.8	16.0	341	13	BI120461
20	78.4	15.9	511	10	AV527020
21	78.4	15.9	528	10	AV551269
22	78.4	15.9	622	9	AU239852
23	78	15.9	322	12	BF428639
24	76.4	15.5	783	10	BE641140
c	74.8	15.2	840	17	AG043467
26	74.6	15.2	724	17	AZ197686
c	74.4	15.1	939	17	AG043613
28	72	14.6	724	17	AZ197686
c	71.4	14.5	843	17	CNS00CS1
30	71	14.4	1212	14	BQ069498
31	70.8	14.4	1192	13	BM556419
c	70.2	14.3	1160	17	AG043473
33	70	14.2	624	9	AI727969
c	69.8	14.2	1168	14	BM809732
35	69	14.0	421	17	BH254057
36	69	14.0	423	14	BQ094569
37	69	14.0	505	12	BG651294
38	69	14.0	551	13	BI972349
39	69	14.0	553	14	BM731926
40	69	14.0	584	13	BM524744
41	69	14.0	609	12	BG238817
42	69	14.0	627	10	AW760124
43	68.4	13.9	1101	17	CNS00397
c	67.6	13.7	729	17	CNS03JVO
45	67.4	13.7	945	17	CNS0076K

ALIGNMENTS

RESULT 1
BG103856
LOCUS
DEFINITION RH122_37_F05.bl_A003 Rhizome2 (RH122) Sorghum propinquum cDNA, mRNA
505 bp mRNA linear EST 30-JAN-2001
sequence.
ACCESSION BG103856
VERSION BG103856.1 GI:12618689
KEYWORDS EST.
SOURCE Sorghum propinquum.
ORGANISM Sorghum propinquum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE 1 (bases 1 to 505)
AUTHORS Cordonnier-Pratt,M.-M., Gingle,A., Paterson,A., Sudman,M. and Pratt ,L.H.
TITLE An EST database from Sorghum: Sorghum propinquum rhizomes
JOURNAL Unpublished (2000)
COMMENT Contact: Cordonnier-Pratt MM
Laboratory for Genomics and Bioinformatics
The University of Georgia, Department of Plant Biology
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 583 0210
Email: mmpratt@uga.edu
Seq primer: JEN REV
High quality sequence stop: 495
POLYA-No. Location/Qualifiers

AU093884 AU093884
BE418829 SCL081.HO
BF631396 HVSME001
BQ467759 HS04M21r
BE498543 WHE0971_F
BI959675 HVSME002
BG414965 HVSME000
BQ801093 WHE2810_C
BQ608457 BRY 4362
BE443955 WHE1123_D
BI157536 ENTRL88TR
BI120461 F015P69Y
AV527020 AV527020
AV551269 AV551269
AU239852 AU239852
BF428639 WHE1411_E
BE641140 Cr12_2_GO
AG043467 Pan trogl
AZ197686 SP_1035_B
AG043613 Pan trogl
AZ197686 SP_1035_B
AL059666 Drosophil
BQ069498 AGENCOURT
BM556419 AGENCOURT
AG043473 Pan trogl
AI727969 BNLGH1984
BM809732 AGENCOURT
BH254057 SALK_0159
BQ094569 san4809
BG651294 sad99g06
BI972349 sag90c12
BM731926 sal86e03
BM524744 sal19c08
BG238817 sab53e03
AW760124 sl58405_Y
AL063912 Drosophil
AL247391 Tetradon
AL066880 Drosophil

```

source          1. .505
/organism="Sorghum propinquum"
/db_xref="taxon:132711"
/clone_lib="Rhizome2 (RHIZ2)"
/notes="Organ: Rhizomes; Vector: pBluescript II from Lambda
zap II; Site:1: XhoI; Site:2: EcoRI; The library was made
from poly-A RNA in the cloning vector lambda ZAP II.
Clones to be sequenced were prepared by mass excision."
BASE COUNT    104 a 131 c 170 g 100 t
ORIGIN

Query Match      60.4%; Score 297; DB 12; Length 505;
Best Local Similarity 89.4%; Pred. No. 4.5e-50;
Matches 329; Conservative 0; Mismatches 38; Indels 1; Gaps 1;

QY 125 GCCTCCGCAAAACCTTAACCCCGGCTCTCGCCCTCCCGGCTCCGCGCCCTCGCCGCTG 184
Db 9 GCCTCCCAAAACCTTAACCCCGGCTCTCGCCCTCCGCGCTCCGAGCGCATCGCGCTG 68

QY 185 TCGTCCGCGGTCTTAGCGCGGAGTGGCGCCAGCCGATGAGTACGGTTAGGCGCTCGTTC 244
Db 69 CGTCCGCGGCTTAACCGCGCGCGCGCGCGCGCGGATGTGTTCGTACGGGCTTCGTCC 128

QY 245 GCCGCTGGCGCGCGCGGAGAGTGGCGGAGTGGGAGGAGTGGTTCAGAACTTCTCCGC 304
Db 129 CGCGCTGGTGGCGGCTGGGAGCGCGGCGGAGTGGGAGGAGTGGCGACGAACGGCTG 188

QY 305 GTGGAGGTGGCGCAACGGCGCGGACGCTGCTGGGGAGGTGCTCANGAAGTACTTCCGC 364
Db 189 GTGGAGGTGGCGAGCGGCGGCGACACGCGCGGAGGTGCTCAGGAAGTACTTCCGC 248

QY 365 CAGCGGTTGAGATCATGCACAAAGAGGACACACAGTCTCTGTACAAATTCAGATAGA 424
Db 249 CAGCGGTTGAGATCATGCACAAAGAGGACACACAGTCTGTACAAATTCAGATAG-AGA 307

QY 425 AGCAGAAGCAATGGTGCTAGTTATCTGAAGAGCTTCCCTCAAGCCAAATTTGG 484
Db 308 AGCAGAAGCAATGGTGCTAGTTATCTGAAGAGCTTCCCTCAAGCCAAATTTGG 367

QY 485 NGAAGAGA 492
Db 368 TGAGGAGA 375

RESULT 2
BG053327      533 bp mRNA linear EST 25-JAN-2001
LOCUS         RHIZ2_26_A12_b1_A003 Rhizome2 (RHIZ2) Sorghum propinquum cDNA, mRNA
DEFINITION   sequence.
VERSION      BG053327.1 GI:12508897
KEYWORDS     EST.
SOURCE       Sorghum propinquum.
ORGANISM     Sorghum propinquum.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Sorghum.
1 (bases 1 to 533)
Cordonnier-Pratt,M.-M., Gingle,A., Paterson,A., Sudman,M. and Pratt
,L.H
An EST database from Sorghum: Sorghum propinquum rhizomes
Unpublished (2000)
Contact: Cordonnier-Pratt MM
Laboratory for Genomics and Bioinformatics
The University of Georgia, Department of Plant Biology
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 583 0210
Email: mmpratt@uga.edu
Seq primer: JEN REV
High quality sequence stop: 507
POLYA=No.
Location/Qualifiers
FEATURES

```

```

source          1. .533
/organism="Sorghum propinquum"
/db_xref="taxon:132711"
/clone_lib="Rhizome2 (RHIZ2)"
/notes="Organ: Rhizomes; Vector: pBluescript II from Lambda
zap II; Site:1: XhoI; Site:2: EcoRI; The library was made
from poly-A RNA in the cloning vector lambda ZAP II.
Clones to be sequenced were prepared by mass excision."
BASE COUNT    111 a 137 c 176 g 109 t
ORIGIN

Query Match      60.4%; Score 297; DB 12; Length 533;
Best Local Similarity 89.4%; Pred. No. 4.6e-50;
Matches 329; Conservative 0; Mismatches 38; Indels 1; Gaps 1;

QY 125 GCCTCCGCAAAACCTTAACCCCGGCTCTCGCCCTCCCGGCTCCGCGCCCTCGCCGCTG 184
Db 9 GCCTCCCAAAACCTTAACCCCGGCTCTCGCCCTCCGCGCTCCGAGCGCATCGCGCTG 68

QY 185 TCGTCCGCGGTCTTAGCGCGGAGTGGCGCCAGCCGATGAGTACGGTTAGGCGCTCGTTC 244
Db 69 CGTCCGCGGCTTAACCGCGCGCGCGCGCGCGGATGTGTTCGTACGGGCTTCGTCC 128

QY 245 GCCGCTGGCGCGCGCGGAGAGTGGCGGAGTGGGAGGAGTGGGAGGAGTGGGAGGAGTGG 304
Db 129 GCCGCTGGTGGCGGCGGAGTGGGAGCGCGGCGGAGTGGGAGGAGTGGGAGGAGTGGGAG 188

QY 305 GTGGAGGTGGCGCAACGGCGCGGACGCTGCTGGGGAGGTGCTCANGAAGTACTTCCGC 364
Db 189 GTGGAGGTGGCGAGCGGCGGCGACGCGCGGAGGTGCTCAGGAAGTACTTCCGC 248

QY 365 CAGCGGTTGAGATCATGCACAAAGAGGACACACAGTCTCTGTACAAATTCAGATAGA 424
Db 249 CAGCGGTTGAGATCATGCACAAAGAGGACACACAGTCTGTACAAATTCAGATAG-AGA 307

QY 425 AGCAGAAGCAATGGTGCTAGTTATCTGAAGAGCTTCCCTCAAGCCAAATTTGG 484
Db 308 AGCAGAAGCAATGGTGCTAGTTATCTGAAGAGCTTCCCTCAAGCCAAATTTGG 367

QY 485 NGAAGAGA 492
Db 368 TGAGGAGA 375

RESULT 3
BG240879      444 bp mRNA linear EST 03-MAY-2002
LOCUS         TaE05011G09R TaE05 Triticum aestivum cDNA clone TaE05011G09R, mRNA
DEFINITION   sequence.
ACCESSION    BG240879
VERSION      BG240879.1 GI:20436755
KEYWORDS     EST.
SOURCE       bread wheat.
ORGANISM     Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
; Triticeae; Triticum.
1 (bases 1 to 444)
Cloutier,S.
Wheat functional genomics - Glenlea developing seeds cDNA libraries
Unpublished (2002)
Contact: Dr. Sylvie Cloutier
Cereal Research Centre, Agriculture and Agri-food Canada
195 Dafoe Rd, Winnipeg, MB, Canada R3T 2M9
Tel: (204) 983-2340
Fax: (204) 983-4604
Email: scloutier@agr.ca
was cloned directionally, not all sequences generated with reverse
primer were from the 5' end (same with forward primer and 3' end).
Average insert size is >2.0 kb
Plate: 011 row: G column: 09
Seq primer: M13 Reverse.
Location/Qualifiers
FEATURES

```

source 1. .444
/organism="Triticum aestivum"
/cultivar="Glenlea"
/db_xref="taxon:4565"
/clone="TaE05011G09R"
/clone_lib="TaE05"
/tissue_type="developing seeds"
/dev_stage="5 days after anthesis"
/lab_host="E. coli DH10B"
/note="Vector: pSPORT-p (Invitrogen Technologies); Site_1:
NotI; Site_2: MluI; mRNA obtained from wheat seeds of
cultivar Glenlea 5 days post-anthesis"
76 a 141 c 146 g 81 t

BASE COUNT 76 a 141 c 146 g 81 t

ORIGIN

Query Match 36.0%; Score 177.2; DB 14; Length 444;
Best Local Similarity 69.1%; Pred. No. 5.5e-26;
Matches 288; Conservative 0; Mismatches 116; Indels 13; Gaps 3;

7y 88 CTCACCCACTCGGCCACCTCGCCCTTCCCGCGCTCGCTCGCGCGTCTT----- 198
Db 25 CTCACCCACTCGGCCACCTTCCCTTCCCATCCGCGCTTCCGAAACCTAGCCTCCG 84

QY 148 CTCCTCGCTCGCTCGCCCTCCCTCTCTGCGCAACCTTCTCTCCGCGCGCAGTTCGGC 144

Db 85 TCACCACTCGCTCGCCCTCCCTCTCTGCGCAACCTTCTCTCCGCGCGCAGTTCGGC 144

QY 199 GAGCGGAGTGGCGCCAGCGATGATGAGGCTTGGCGCTCGCTCGCGCGTGGCGGC 258

Db 145 GTCGGGATAGCGGGCGTGGATGGCTCGGTTCGAGCTTCCCTCTGAGCGGAGG 204

QY 259 CGGCGGAGAGTGGCGCAG---TGGGGAGTTGGCGAGCGGCTGGTGGAGTGGC 315

Db 205 CTGGCGGTGGCTCGCGCGGTGAAGAGGGGTGGAGATGGAGCGCTGGTGGCGTGGC 264

QY 316 GCAACGGCGCGGCGACCTCTCGGGAGTGTCTCANGAAGTACTTCCGCGCGGTTGA 375

Db 265 GCAGCGCGCGGATGCGCGGGGAGTGTCTCANGAAGTACTTCCGCGCGGTTGA 324

QY 376 GATCATCGAAGAGAGACACAGTCTCTTACAATTGCGATAGAGAAGCAGAAGAAG 435

Db 325 GATCATCGAAGAGAGACACAGTCCGGTTCAGCATCGCTGATAG-AGAAGCAGACGAAG 383

QY 436 CAATGGNGTCAGTATATCAAGAGCTTCCCTACTCAAGCCAAATTTGGNGAAGAGA 492

Db 384 CAATGAGCTCAGTACTGAATAGCTTCTCTACTCATCTGTTTTCGCGGAGGAGA 440

RESULT 4
BI778441

LOCUS BI778441 588 bp mRNA linear EST 23-JUL-2002

DEFINITION Ebro07_SQ003_F02_R root, 3 week, reduced light, cv Optic, Ebro07

ACCESSION BI778441

VERSION BI778441.2 GI:21950120

KEYWORDS EST.

SOURCE Hordeum vulgare.

ORGANISM Hordeum vulgare.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
; Triticeae; Hordeum.

REFERENCE 1 (bases 1 to 588)
Hedley, P., Liu, H., Caldwell, D., McCallum, N., Mudie, S., Cardie, L.,
Ramsay, L., Machray, G., Marshall, D.F.M. and Waugh, R.
Development of Barley Transcriptome Resources
Unpublished (2001)
On Sep 26, 2001 this sequence version replaced gi:15781333.
Contact: Waugh R, Marshall DF
Genome Dynamics/Computational Biology
Scottish Crop Research Institute
Invergowrie, Dundee DD2 5DA, Scotland, UK
Tel: 00 44 1382 562731
Fax: 00 44 1382 562426

Email: est@scri.sari.ac.uk
All sequence has a Phred quality score of 20 or over
Seq primer: M13 reverse.

FEATURES
Location/Qualifiers
source 1. .588
/organism="Hordeum vulgare"
/cultivar="Optic"
/db_xref="taxon:4513"
/clone="Ebro07_SQ003_F02"
/clone_lib="root, 3 week, reduced light, cv Optic, Ebro07"
/tissue_type="root"
/dev_stage="3 week"
/lab_host="DH10B"
/note="Vector: pSPORT1; Site_1: Sal I; Site_2: Not I;
Non-normalised library, directionally cloned into pSPORT1.
Derived from shoot and root material of 3 week old
etiolated barley plants. Developed as part of the barley
transcriptome resources of BBSRC/SERAD funded cereal IGF
(Investigating Gene Function) project."

BASE COUNT 98 a 200 c 183 g 106 t 1 others

ORIGIN

Query Match 35.8%; Score 176; DB 13; Length 588;
Best Local Similarity 66.7%; Pred. No. 9.7e-26;
Matches 318; Conservative 0; Mismatches 142; Indels 17; Gaps 4;

QY 28 CATTTAAACCTTGTCTCCGACACACCGCGGCGACTCTCTCCCAATGCTCTCTCTCTCTC 87

Db 104 CATTTAAACCTGCGCTCTGCGCGCGCTTTCGCCCTCCGAGATGCTTCTCC---CAAC 159

QY 88 CTCACCCACTCGCGCGACCTCGCCCTTCCCGCGCTCGCTCGCGCAACCTTAACCCCG 147

Db 160 CTCACCCACTCCACCCACCTTCCCTTCCGCGCGCTTCCGCGCGCTTCCGCGCGCTTCCG 219

QY 148 CTCCTCGCTCTCT-----CGCGCTCGCGCGCTCTCGCGCTGTCTCTCGCGCTCTT 198

Db 220 TCACCACTCTCGCT 279

QY 199 GAGCGGAGTGGCGCGCGCGATGATGAGTGGCTTGGCGCTGTCTCGCGCTGGCGCGC 258

Db 280 GTGCGGATAGCGCGCGCTTGGACGAGCTCGGTTCGAGCTTCTGAGCTTCTGAGCGCGGG 339

QY 259 CGCGCGGAGAGTTCGCGCAG---TGGGGAGTTGGCGAGCGGCTGGTGGAGTGGC 315

Db 340 CTGCGCGTGGCTGCGCGTGAAGGAGGGGTGGAGATGGAGCGGCTGGTGGCGGTGGC 399

QY 316 GCAACGGCGCGCGAGCTGTCTGCGGAGTGTCTCANGAAGTACTTCCGCGCGCGGTTGA 375

Db 400 GCAGAGCGCGCGCGATGATGATGCGCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 459

QY 376 GATCATCGAAGAGAGTACAGTCTCTTACAATTGCGATAGAGAAGCAGAAGAAG 435

Db 460 GATAATCGAAGAGAGGAGGACACAGTCCGCTCAGCTGCTGATAG-AGNAGCAGAAGAGG 518

QY 436 CAATGGNGTCAGTATATCAAGAGCTTCCCTACTCAAGCCAAATTTGGNGAAGAGA 492

Db 519 CAATGAGCTCAGTACTGAAGAGCTTCTCTACTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 575

RESULT 5
BI951067

LOCUS BI951067 863 bp mRNA linear EST 19-OCT-2001

DEFINITION HVSME10024E07f Hordeum vulgare spike EST library HVCNDA0012

ACCESSION BI951067

VERSION BI951067.1 GI:16293708

KEYWORDS EST.

SOURCE Hordeum vulgare.

ORGANISM Hordeum vulgare.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
; Triticeae; Hordeum.

QY	77	TCCTCTTCTCTCTCACACCACTCGGCCCACTCGCCCTTCCCGCGCTCGCCCGCAAC	136
Db	179	CCC-----CAACCTCCACCATCCACCCACCTTCCCTTCCCCACCCCGCTTCACAAA	234
QY	137	CCTAACCCCGCTCTCGGCTCTCT-----CCGCCCTCCGCGCGCTCGCCCGTGTCTG	187
Db	235	CCTAGCTTCCGTCACACCTCGCTCGCTCTCTCTCCCTGGCAAGCGCTCTCTCGCGCG	294
QY	188	TCGCGGCTTTGAGCGCAGTGGCGCCGACCGGATGAGTACGGTTAGGCGCTCGTTCGCC	247
Db	295	GCGGTCGCGGCTGCGGATACGCGGCCGTTGGACGAGCTCGGTTTCGAGCGCTTCCT	354
QY	248	CTGGGCGCGCGCGGAGAGCTCGGCAG---TGGGGGAGTTGGCGACGAGCGCGCTG	304
Db	355	GAGCGCGGGGCTGGCGGCTGCTCGGCTGTAAGAGAGGGGTGGAGATGGAGCGCTG	414
QY	305	GTGAGGTGGCGCAACGGCGCGGACCTCGTGGGGAGGTGCTCANGAAGTACTTCGCG	364
7b	415	GTGGCGGTGGCGCATAGCGCATCGGATCGCGCGGGGAGGTGCTCATGAAGTACTTCAG	474
QY	365	CAGCGGTTGAGATCATCGACAAGAGCACACAGTCTCTGTACAAATTCAGATAGAGA	424
Db	475	CAGCGTCCAGATAATCGACAAGAGGACACAGTCCCGTACAGATCCGTGATAG-AGA	533
QY	125	AGCAGAAGCAATGGNGTCAGTTACTGAAGAGC TCCTCACTCAAGCAATTTTGG	484
Db	534	AGCAGAAGCAATGACTTCAGTCATACTGAAGAGCTTCTCTACTCATGCTGTTTCGG	593
QY	485	NGAAG 489	
Db	594	CGAGG 598	
RESULT 7			
LOCUS	AU093884	483 bp	linear EST 03-APR-2002
DEFINITION	AU093884	Rice panicle at flowering stage	Oryza sativa (japonica cultivar-group) cDNA clone EI081, mRNA sequence.
ACCESSION	AU093884		
VERSION	AU093884.1	GI:8856566	
KEYWORDS			EST.
SOURCE			Oryza sativa (japonica cultivar-group).
ORGANISM			Oryza sativa (japonica cultivar-group)
REFERENCE			Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoidae; Oryzaceae; Oryza.
AUTHORS			Sasaki,T. and Yamamoto,K.
TITLE			Rice cDNA from panicle at flowering stage (2000)
JOURNAL			Unpublished (2000)
COMMENT			Contact: Takuji Sasaki National Institute of Agrobiological Resources Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki 305-8602, Japan Tel: 81-298-38-7441 Fax: 81-298-38-7468 Email: tsasaki@abr.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/ PROJECT = 'RGP'.
FEATURES			Location/Qualifiers
source		1..483	/organism="Oryza sativa (japonica cultivar-group)" /cultivar="Nipponbare" /db_xref="taxon:39947" /clone="EI081" /clone_lib="Rice panicle at flowering stage" /dev_stage="flowering stage" /note="Organ: panicle; Rice cDNA from panicle at flowering stage"
BASE COUNT	87 a	138 c	172 g 84 t 2 others
ORIGIN			
Query Match		34.2%;	Score 168.4; DB 9; Length 483;


```
/clone_lib="ITEC SCL Wheat Leaf Library"
/tissue_type="young leaf"
/dev_stage="after 24 hour challenge with LR-AVR1"
/notes="Vector: Lambda ZAP; 1.0 Kbp average insert size."

BASE COUNT      194 a   243 c   242 g   171 t   14 others
ORIGIN

Query Match      34.2%; Score 168.4; DB 10; Length 864;
Best Local Similarity 67.3%; Pred. No. 3.3e-24;
Matches 329; Conservative 0; Mismatches 141; Indels 19; Gaps 6;

QY 17 CTCGCCGCCGACATTTAAACCTTGTCTCCGACAAACGCGCGGCTCTCCCAATGCTC 76
Db 115 CGCGCGACACATTTAAACCTGCGCGCTGCGCGCGC--TTTTTCTCCAGATGCTT 172
QY 77 TCCTCTTCTCTCCACCCACTGGCCACTCGCCCTTCCCGGCTCGCTCCGCAAC 136
Db 173 CCGCCAATCTGCCACCATCC---CGCTCACCTTTCCCTTCCCAACCGCGCTTCACAAA 229
QY 137 CCTAACCCCGCTCTCGCCCTCCCTCCGCGCGCGCTCGCCCTGCTCGTCGGCGGTC 196
Db 230 CCTAGCTCTCGTCCACACCTCGCTCGCCCTCCTCTCTGCAACCTTCTCTCCGCGGG 289
QY 197 TTGAGC-----CGAGTGGCGGCCAGCCGATGAGTACGGTTAGGGCCTCGTTCCGC 247
Db 290 GCGGGTGGCGGTGTGGGATAGCGGGCGTTGGATGGGCTCGGTTGAGCTCGCCCTCT 349
QY 248 GCTGGGCGCGCGCGGAGAGCTGCGGAG---TGGGGAGTTGCGAGCGGAGCGGTC 304
Db 350 GAGCGGGGGCTGGCGGTGCTGCGCGGTGAAGAGGGGTGGAGATGAGCGGCTG 409
QY 305 GTGAGGTGGCGCAACGGCGCGGCGCTGCTGGGGAGTGTCTCANGAAGTACTTCCGC 364
Db 410 GTGGCGTGGCCCANAGCGCGCGGATGCGCGGGGCGTGTGCTCAGGAAGTACTTCAG 469
QY 365 CACGGGTTGAGATCATGCAGAAAGAGGACACAGTCTCTGTACATTCAGATAGAGA 424
Db 470 CACGGCTTCAGATCATGCAGAAAGAGGACACAGTCTGCTCAGATCGCTGATAG-AGA 528
QY 425 AGCAGAAGAGCAATGGNGTCTAGTTACTAGTGAAG-AGCTTCTCTACTCAAGCAATTTG 483
Db 529 AGCAGGAAGCAATCACGTCACTAGTATCTGAAGAGCTTCTCTACTCATGCTGTTTCG 588
QY 484 GCAAGAGA 492
Db 589 GCGAAGANA 597

"RESULT 9
"631396
LOCUS
DEFINITION
BF631396 766 bp mRNA linear EST 22-OCT-2001
HVSMEB0015N03f Hordeum vulgare seedling shoot EST library
HVCNDA0002 (Dehydration stress) Hordeum vulgare cDNA clone
HVSMEB0015N03f, mRNA sequence.
ACCESSION
BF631396.2 GI:13092082
VERSION
EST.
KEYWORDS
Hordeum vulgare.
SOURCE
Hordeum vulgare.
ORGANISM
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Lilioidae; Poales; Poaceae; Pooideae
; Triticeae; Hordeum.
1 (bases 1 to 766)
REFERENCE
Wing,R., Close,T.J., Kleinohs,A., Wise,R., Begum,D., Frisch,D., Yu
,Y., Henry,D., Palmer,M., Rambo,T., Simmons,J., Choi,D.W., Fenton
,R.D., Oates,R. and Main,D.
Development of a genetically and physically anchored EST resource
for barley genomics: Morex drought-stressed seedling shoot cDNA
library
Unpublished (2001)
JOURNAL
On Dec 19, 2000 this sequence version replaced gi:11895554.
COMMENT
Contact: Wing RA
Clemson University Genomics Institute
```

```
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Total hg bases = 565
Seq primer: AATTAACTCTCACTAAAGGG
High quality sequence stop: 740.
Location/Qualifiers
1..766
/organism="Hordeum vulgare"
/cultivar="Morex"
/db_xref="taxon:4513"
/clone="HVSMEB0015N03f"
/clone_lib="Hordeum vulgare seedling shoot EST library
HVCNDA0002 (Dehydration stress)"
/tissue_type="Seedling shoot"
/lab_host="TJCL21"
/notes="Vector: lambdaZAP; Site_1: EcoRI; Site_2: XhoI;
Seeds were surface sterilized then germinated under axenic
conditions in the dark at room temperature on filter paper
with water, nystatin and cefotaxime in covered
crystallization dishes. Five-day old seedlings were
incubated at 90% RH for 24 hr. Shoots were then harvested,
total RNA was prepared, poly(A) RNA was purified, one
primary unamplified cDNA library was made, 600000 pfu were
in vivo excised to give phagescript SK(-) cDNA phagemids.
These steps were performed in the TJ Close laboratory at
the University of California, Riverside (Choi, Close,
Fenton). Phagemids were plated and picked at the Clemson
University Genomics Institute (CUGI) (Begum, Palmer,
Frisch, Atkins and Wing). plasmid DNA preparations, DNA
sequencing and sequence analysis were performed at CUGI
(Wing, Yu, Frisch, Henry, Simmons, Oates, Rambo, Main).
The sequence has been trimmed to remove vector sequence
and contains a minimum of 100 bases of phred value 20 or
above. For more details on library preparation and
sequence analysis see
http://www.genome.clemson.edu/projects/barley. To order
this clone see http://www.genome.clemson.edu/orders/Also
see clone TJ, Wing R, Kleinohs A, Wise R (2001)
Genetically and physically anchored EST resources for
barley genomics. Barley Genetics Newsletter 31:29-30.
(http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html)"

BASE COUNT      165 a   177 c   242 g   182 t
ORIGIN

Query Match      32.7%; Score 161; DB 12; Length 766;
Best Local Similarity 68.9%; Pred. No. 1e-22;
Matches 248; Conservative 0; Mismatches 108; Indels 4; Gaps 2;

QY 136 CCTAACCCCGCTCTCGCCCTCTCCGCTCGCGCGCGCTCGCGCGGT 195
Db 30 CGTCAACACCTCCGCTCGCTCTCGCGCGCGCTCGCGCGGT 89
QY 196 CTTGAGCGCGAGTGGCGGCCAGCCGATGATACGGTTAGGGCCTGTTTCGGCGTGGGGC 255
Db 90 GCGGTGCGGGATAGCGGGCGCTTGGACGAGCTCGGTTCAGCCTCGCCTCTGAGGCGGG 149
QY 256 GCGCGCGCGGACAGCTGCGGCGCT---GGGGAGTTGCGGAGGAGCGGCTGTTGGAGGT 312
Db 150 GGGCTGGCGGCTGCGGCGCTGTTGAGGAGGGGTGGAGATGAGCGGCTGTTGGCGGT 209
QY 313 GCGCAACGGCGCGGCGGAGCTGCTGGGGAGGTGCTCANGAAGTACTTCCGCCAGCGGCT 372
Db 210 GCGCAGAGCGCAGCGGATGCGCGGGGAGGTGCTCAGGAAGTACTTTCAGGCGCGCTT 269
QY 373 TGAGATCATCACAAGAGGACACAGTCTCTTACAAATTCAGATAGAGAAGCAGAAAG 432
Db 270 CGAGATAATCACAAGAGGAGGACACAGTCCCGTCCAGCTCGCTGATAG-AGAAGCAGAA 328
QY 433 AAGCAATGCGTCACTTATCTGAAGAGCTTCCCTACTCAAGCCCAATTTTGGNAGAGA 492
1 11111 1111 1111111111 111111 11 111111 11 1111111111
```

```
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Total hg bases = 565
Seq primer: AATTAACTCTCACTAAAGGG
High quality sequence stop: 740.
Location/Qualifiers
1..766
/organism="Hordeum vulgare"
/cultivar="Morex"
/db_xref="taxon:4513"
/clone="HVSMEB0015N03f"
/clone_lib="Hordeum vulgare seedling shoot EST library
HVCNDA0002 (Dehydration stress)"
/tissue_type="Seedling shoot"
/lab_host="TJCL21"
/notes="Vector: lambdaZAP; Site_1: EcoRI; Site_2: XhoI;
Seeds were surface sterilized then germinated under axenic
conditions in the dark at room temperature on filter paper
with water, nystatin and cefotaxime in covered
crystallization dishes. Five-day old seedlings were
incubated at 90% RH for 24 hr. Shoots were then harvested,
total RNA was prepared, poly(A) RNA was purified, one
primary unamplified cDNA library was made, 600000 pfu were
in vivo excised to give phagescript SK(-) cDNA phagemids.
These steps were performed in the TJ Close laboratory at
the University of California, Riverside (Choi, Close,
Fenton). Phagemids were plated and picked at the Clemson
University Genomics Institute (CUGI) (Begum, Palmer,
Frisch, Atkins and Wing). plasmid DNA preparations, DNA
sequencing and sequence analysis were performed at CUGI
(Wing, Yu, Frisch, Henry, Simmons, Oates, Rambo, Main).
The sequence has been trimmed to remove vector sequence
and contains a minimum of 100 bases of phred value 20 or
above. For more details on library preparation and
sequence analysis see
http://www.genome.clemson.edu/projects/barley. To order
this clone see http://www.genome.clemson.edu/orders/Also
see clone TJ, Wing R, Kleinohs A, Wise R (2001)
Genetically and physically anchored EST resources for
barley genomics. Barley Genetics Newsletter 31:29-30.
(http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html)"
```

FEATURES
source

```

Db 329 AGCAATGACTTCAGTCATAGAGCTTTCTCTACTCATGCTGTTTTTCGGCGGAGGAGA 388
RESULT 10
B0467759
LOCUS HS04M21r HS Hordeum vulgare cDNA clone HS04M21 5-PRIME, mRNA
DEFINITION B0467759
ACCESSION B0467759.1 GI:21275541
VERSION B0467759
KEYWORDS EST.
SOURCE Hordeum vulgare.
ORGANISM Hordeum vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae
1 (bases 1 to 443)
Zhang, H., Potokins, E., Michalek, W., Weschke, W., Stein, N. and Graner
, A.
Barley ESTs from germinating seeds
Unpublished (2002)
Contact: Stein Nils
Molecular Markers Group, Department Genbank
Institute of Plant Genetics and Crop Plant Research (IPK)
Corrensstr. 3, 06466, Gatersleben, Germany
Tel: 039482-5522
Fax: 039482-5595
Email: stein@ipk-gatersleben.de
Insert Length: 443 Std Error: 0.00
Plate: 4 Row: M Column: 21
Seq primer: M3rev.
Location/Qualifiers
1. .443
/organism="Hordeum vulgare"
/cultivar="Barke"
/db_xref="taxon:4513"
/clone="HS04M21"
/tissue="embryo + scutellum"
/dev_stage="0-16 hours after imbibition"
/lab_host="XL10-Gold"
/notes="Vector: pBluescript SK+; Site.1: EcoRI (5'-end of
cDNA); Site.2: XhoI (3'-end of cDNA); Due to a cloning
artefact caused by the kit, in most cases the EcoRI site
is NOT present, as well as the EcoRI adapter used for
cloning. To excise the insert, restriction sites upstream
EcoRI should be used (e.g. BamHI, Sali, PstI). NOTE: Also
due to the cloning system used Blue/white selection for
recombinants is not 100% reliable."
JASE COUNT 93 a 94 c 161 g 95 t
ORIGIN
Query Match 31.5%; Score 155.2; DB 14; Length 443;
Best Local Similarity 74.6%; Pred. No. 1.4e-21;
Matches 220; Conservative 0; Mismatches 71; Indels 4; Gaps 2;
QY 201 GCGCGAGTGGCGCCAGCGATGAGTACGTTAGGCGCTTCGCGCTGGCGCGCGC 260
Db 3 GCACGAGCGGGCGCTTGGAGAGCTCGGTTGCGAGCTTCGCTTCTGAGCGCGGGGCT 62
QY 261 GCGCGAGTGGCGCGCAG---TGGGGAGTTGGCGACGAGCGGCTGTGGAGGTGGCGC 317
Db 63 GGGCGGTGGCTGGCGCTGGTGAAGAGGGGGTGGAGATGAGCGGCTGTGGCGGTGGCGC 122
QY 318 AACGGCGCGGACGCTGTGGGGAGGTGCTCANGAAGTACTTTCGCCAGCGGGTTGAGA 377
Db 123 AGAGCGCAGCGATGGCGGGGGAGGTGCTCAGGAAGTACTTTCAGCGAGCGCTTCGAGA 182
QY 378 TCATCACAAAGAGACACACACTCTGTTTACAAATTGCAGATAGAGACGACGACGACA 437
Db 183 TAATCACAAGAGGAGGACACACTCCGTCACGATCGCTGATAG-AGAAGCAGAAGGCA 241
QY 438 ATGGNGTCAGTTATCTAGAGAGCTTCCTCTACTCAAGCAATTTTGGNGAGAGA 492

```

```

Db 242 ATGACTTCAGTCATAGAGCTTTCTCTACTCATGCTGTTTTTCGGCGGAGGAGA 296
RESULT 11
BE498543
LOCUS WHE0971_F05_K092S Wheat pre-anthesis spike cDNA library Triticum
DEFINITION WHE0971_F05_K092S Wheat pre-anthesis spike cDNA library Triticum
ACCESSION BE498543
VERSION BE498543
KEYWORDS EST.
SOURCE bread wheat.
ORGANISM Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae
1 (bases 1 to 404)
Anderson, O.D., Chao, S., Choi, D.W., Close, T.J., Fenton, R.D., Han
, P.S., Hsia, C.C., Kang, Y., Lazo, G.R., Miller, R., Rausch, C.J.,
Seaton, C.L. and Fong, J.C.
The structure and function of the expressed portion of the wheat
genomes - pre-anthesis spike cDNA library
Unpublished (2000)
Contact: Olin Anderson
US Department of Agriculture, Agriculture Research Service, Pacific
West Area, Western Regional Research Center
800 Buchanan Street, Albany, CA 94710, USA
Tel: 5105595773
Fax: 5105595818
Email: oanderson@pw.usda.gov
Sequence have been trimmed to remove vector sequence and low
quality sequence with phred score less than 20
: q primer: Stratagene SK primer.
Location/Qualifiers
1. .404
/organism="Triticum aestivum"
/cultivar="Chinese Spring"
/db_xref="taxon:4565"
/clone="WHE0971_F05_K09"
/tissue="Wheat pre-anthesis spike cDNA library"
/dev_stage="Adult plant"
/lab_host="E. coli SOLR"
/notes="Vector: Lambda Uni-ZAP XR, excised phagemid;
Site.1: EcoRI; Site.2: XhoI; Plants were grown in the
greenhouse. Whole spike with awns trimmed, white, green
and yellow anther were collected and total RNA, and
poly(A) RNA were prepared, a cDNA library was made, and
the cDNA clones were in vivo excised to give pBluescript
phagemids in the TJ Close lab (Choi, Close, Fenton) at
the University of California, Riverside. Plasmid DNA
preparations and DNA sequencing were performed in the OD
Anderson lab (all other authors)."
BASE COUNT 91 a 83 c 138 g 89 t 3 others
ORIGIN
Query Match 30.2%; Score 148.8; DB 10; Length 404;
Best Local Similarity 81.7%; Pred. No. 2.8e-20;
Matches 183; Conservative 0; Mismatches 40; Indels 1; Gaps 1;
QY 269 GCTGGCGAGTGGGGAGTTGGCGACGAGCGGTTGGAGGTGGCGCAACGGCGCGC 328
Db 21 GCGCGGGTAAAGAGGGGGTGACATGAGCGGCTGTGGCGGTGGCGCANAGCGCGC 80
QY 329 GACGCTGCTGGGAGGTGCTCANGAAGTACTTCGCCAGCGGTTGAGATCATGACAAA 388
Db 81 GATCGCGCGGGAGGTGCTCANGAAGTACTTCAGGAGCGGCTTCGAGATCATGACAAA 140
QY 389 GAGGACACAGTCTCTGTTTACAAATTCAGATAGAAAGACGAGAAGCAATGGNGTCAGT 448
Db 141 GAGGACACAGTCCCGTCACGATCGCTGATAG-AGAGCAGAGAAGCAATGACTCAGT 199

```



```

/clone_lib="Hordeum vulgare testa/pericarp EST library
HVCNA0013 (normal)"
/tissue_type="testa/pericarp"
/lab_host="TJC121"
/notes="Vector: lambdaZAP; Site_1: EcoRI; Site_2: XhoI;
Plants were raised from seeds in a Controlled Environments
growth chamber maintained in continuous light at 18oC, and
testa and pericarp were dissected from developing kernels
at Washington State University, Pullman, WA (Kannangara,
von Wettstein). Total RNA was prepared, poly(A) RNA was
purified, one cDNA library was made, and 1 million pfu
were in vivo excised to give pBluescript SK(-) cDNA
phagemids in the T3 Close lab at the University of
California, Riverside (Akhunov, Chin, Choi, Close, Fenton,
Kianian, Otto, Simons, Zhang). Phagemids were plated and
picked at the Clemson University Genomics Institute (CUGI)
(Begum, Palmer, Frisch, Atkins and Wing). Plasmid DNA
preparations, DNA sequencing and sequence analysis were
performed at CUGI (Wing, Yu, Frisch, Henry, Simmons, Oates
, Rambo, Main). The sequence has been trimmed to remove
vector sequence and contains a minimum of 100 bases of
phred value 20 or above. For more details on library
preparation and sequence analysis see
http://www.genome.clemson.edu/projects/barley. To order
this clone see http://www.genome.clemson.edu/orders/Also
see clone T3, Wing R, Kleinhofs A, Wise R (2001)
Genetically and physically anchored EST resources for
barley genomics, Barley Genetics Newsletter 31:29-30.
(http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html)"

```

```

BASE COUNT 153 a 127 c 193 g 157 t 1 others
ORIGIN
Query Match 28.8%; Score 141.8; DB 12; Length 631;
Best Local Similarity 79.5%; Pred. No. 7.3e-19;
Matches 178; Conservative 0; Mismatches 45; Indels 1; Gaps 1;

```

```

QY 269 GCTGCGCAGTGGGAGTTGGCAGCGAGCGTGTGGAGTGGCGCAACGGGCGGCG 328
Db 1 GCGGCTGGTAAGAGGGGTGGAGATGGA3CGGCTGTGGCGGTGGCGCAGCGCAGCG 60
QY 329 GAGCGTCTGGGAGTGTCTCANGAAGTACTTCGCCGACGCGGTTGAGATCATCGACAAA 388
Db 61 GATGCGGCGGGAGGTGCTCAGGAAGTACTTCAGGACGCGCTTCGAGATATCGACAAG 120
QY 389 GAGGACACAGTCTCTTTACAATTCGAGATAGAGAAGCAGAGCAATGGNGTCACT 448
Db 121 GAGGACACAGTCCCGTCAGATCGCTGATAG-AGAAGCAGAAGAGGCAATGACTTCACT 179
QY 449 TATACGAGAGCTTCCCTACTCAACCCAAATTTTCGNGAAGAGA 492
Db 180 CATACTGAAGAGCTTCCCTACTCATGCTGTTCGCGCGAGGAGA 223

```

```

RESULT 14
BQ801093 759 bp mRNA linear EST 30-JUL-2002
LOCUS WHE2810_C03_E06ZS Triticum monococcum vernalized apex cDNA library
DEFINITION Triticum monococcum cDNA clone WHE2810_C03_E06, mRNA sequence.
ACCESSION BQ801093
VERSION BQ801093.1 GI:22016062
KEYWORDS EST.
SOURCE Triticum monococcum.
ORGANISM Triticum monococcum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae
; Triticeae; Triticum.
1 (bases 1 to 759)
Anderson, O. D., Chao, S., Crossman, C., Dubcovsky, J., Echenique, V.,
Lazo, G. R., Pham, J., Rausch, C. J., Stamova, B., Wilson, C. and Woo, J.
The structure and function of the expressed portion of the wheat
genomes - Vernalized apex cDNA library from Triticum monococcum
Unpublished (2002)
Contact: Olin Anderson

```

```

US Department of Agriculture, Agriculture Research Service, Pacific
West Area, Western Regional Research Center
800 Buchanan Street, Albany, CA 94710, USA
Tel.: 5105595773
Fax: 5105595818
Email: oanderson@pw.usda.gov
Sequences have been trimmed to remove vector sequence and low
quality sequence with phred score less than 20
Seq primer: SK primer.

```

FEATURES

```

source
1..759
Location/Qualifiers
1..759
/organism="Triticum monococcum"
/cultivar="G3116"
/db_xref="taxon:4568"
/clone="WHE2810_C03_E06"
/clone_lib="Triticum monococcum vernalized apex cDNA
library"
/tissue_type="Vernalized apex"
/dev_stage="One month old plants"
/lab_host="E. coli XL0LR"
/notes="Vector: Lambda Uni-ZAP XR, excised phagemid;
Site_1: EcoRI; Site_2: XhoI; One-month old plants were
subjected to vernalization treatment by placing them in
the cold room at 6 C. under 15hr light/9hr dark condition.
Total RNA was prepared from apex tissue extracted from
plants with no cold treatment; and from plants with 2-week
, 4-week and 6-week cold treatment separately. Equal
amount of total RNA was pooled from all four samples, a
cDNA library was made using pooled poly(A) RNA and cDNA
clones were in vivo excised at the University of
California, Davis (V. Echenique, B. Stamova, J. Dubcovsky
). Plasmid DNA preparations and DNA sequencing were
performed in the OD Anderson lab (all other authors)."
```

```

BASE COUNT 186 a 153 c 223 g 197 t
ORIGIN
Query Match 24.2%; Score 119.2; DB 14; Length 759;
Best Local Similarity 12.1%; Pred. No. 2.6e-11;
Matches 147; Conservative 0; Mismatches 31; Indels 1; Gaps 1;

```

```

QY 314 GCGCAACGGGCGGAGCTCTCTGGGAGGTGCTCANGAAGTACTTCGCCAGCGGGTT 373
Db 10 GAGGAGAGCGGCGGAGTCTGGCGGGGAGGTCTCAGGAGTACTTCAGGCGGCGCTTC 69
QY 374 GAGATCATCGACAAAGAGACACAGTCTCTGTACAATTCGAGATAGAGAAGCAGAAGA 433
Db 70 GAGATCATCGACAAAGAGAGAGACACAGTCTCTGTACAATTCGAGATAG-AGAAGCAGAGA 128
QY 434 AGCAATGGNGTCACTTATATATCAAGAGCTTCCTACTCAAGCCAAATTTGGNGAAGAGA 492
Db 129 AGCAATGAGTCACTACTGAGAGGCTTCTCTACTCATGCTGTTCGCGCGAGGAGA 187

```

```

RESULT 15
BQ608457 745 bp mRNA linear EST 25-JUN-2002
LOCUS BRY_4362 wheat EST endosperm library Triticum aestivum cDNA 5',
DEFINITION mRNA sequence
ACCESSION BQ608457
VERSION BQ608457.1 GI:21557796
KEYWORDS EST.
SOURCE bread wheat.
ORGANISM Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae
; Triticeae; Triticum.
1 (bases 1 to 745)
Clarke, B., Lambrecht, M. and Rhee, S.
Assessing the utility of Arabidopsis genomic information for
interpreting wheat EST sequences
Unpublished (2002)
Contact: Lambrecht M
The Arabidopsis Information Resource

```

Carnegie Institution of Washington, Dept. of Plant Biology
260 Panama Street, Stanford, CA 94305, USA
Tel: 1 650 325 1521 x 251
Fax: 1 650 325 3748
Email: rhee@acoma.stanford.edu.

FEATURES
source
1. .745
/organism="Triticum aestivum"
/cultivar="Wyuna"
/db_xref="taxon:4565"
/clone_lib="wheat EST endosperm library"
/tissue_type="endosperm"
/dev_stage="developing endosperm tissue 8, 10 and 12 DPA
(days post anthesis)"
/note="Vector: Bluescript II SK(-)"

BASE COUNT 191 a 137 c 217 g 195 t 5 others
ORIGIN

Query Match 23.4%; Score 115.2; DB 14; Length 745;
Best Local Similarity 83.8%; Pred. NO. 1.6e-13;
Matches 140; Conservative 0; Mismatches 26; Indels 1; Gaps 1;

QY 326 GCGGACGCTGCTGGGAGGTGCTCANGAGTACTTCCACAGCGGTTGAGATCATCGAC 385
||||| || |||||||||||| |||||||| ||||| ||||||||||||
Db 1 GCGGATCGCGGGGGAGGTGCTCAAGAGTACTTCAGGACGCTTCGAGATCATCGAC 60
||||| |||||||||||| ||||| ||||| ||||||||||||
QY 386 AAAGAGGACCAAGTCTGTTACAATTGCAGATAGAGAAGCAGAGAAGCAATGGNGTC 445
||||| |||||||||||| ||||| ||||| |||||||||||| ||||| ||
Db 61 AAAGAGGACCAAGTCTGCCGTCAGATCGCTGTAG-AGAAGCAGAGAAGCAATGACCTC 119
||| |||||||||||| |||||||| || ||||| ||||| |||||
QY 446 AGTTATCTCAAGAGCTTCCTACTCAAGCAATTTTGGNGAAGAGA 492
||| |||||||||||| |||||||| || ||||| ||||| |||||
Db 120 AGTCATCTCAAGAGCTTTCTCTACTCATGCTGTTTCGGTGAGGAGA 166

Search completed: November 18, 2002, 20:55:28
Job time : 2122 secs

GenCore version 5.1.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 19, 2002, 00:20:30 : Search time 2718 Seconds
(without alignments)
5268.058 Million cell updates/sec

Title: US-09-686-522C-13
Perfect score: 492
Sequence: 1 agtccgagcgtctctcc.....agccaatttggngaagaga 492

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
c 1	112	22.8	131704	2	AP004313	AP004313 Oryza sat
c 2	112	22.8	161809	2	AP005173	AP005173 Oryza sat
c 3	84.2	17.1	163381	2	AC128230	AC128230 Rattus no
c 4	78	15.9	177168	2	AC107145	AC107145 Rattus no
c 5	76.8	15.6	138467	2	AC111697	AC111697 Rattus no
c 6	74.2	15.1	82815	2	AC062001	AC062001 Homo sapi
c 7	73.2	14.9	158897	2	AC121439	AC121439 Rattus no
c 8	72.2	14.7	68330	2	AC116109	AC116109 Mus muscu
c 9	72.2	14.7	117082	2	AC126076	AC126076 Rattus no
c 10	70.4	14.3	101509	2	AC027353	AC027353 Homo sapi
c 11	69.8	14.2	151358	2	AC125906	AC125906 Rattus no
c 12	68.4	13.9	174605	2	AC121227	AC121227 Rattus no
c 13	68	13.8	174031	2	AC118303	AC118303 Rattus no
c 14	67.4	13.7	68787	2	AC106398	AC106398 Rattus no
c 15	67.4	13.7	130899	2	AC126730	AC126730 Rattus no
c 16	67.4	13.7	165988	2	AC121212	AC121212 Rattus no
c 17	67.2	13.7	185263	2	AC129765	AC129765 Rattus no
c 18	67.2	13.7	188231	2	AC094753	AC094753 Rattus no
c 19	66.8	13.6	179608	2	AC128497	AC128497 Rattus no
c 20	66.8	13.6	209216	2	AC117126	AC117126 Rattus no
c 21	66.6	13.5	99849	2	AC096824	AC096824 Rattus no
c 22	66.6	13.5	117082	2	AC126076	AC126076 Rattus no
c 23	66.6	13.5	118276	2	AP004120	AP004120 Oryza sat
c 24	66.6	13.5	158117	2	AP005111	AP005111 Oryza sat
c 25	66.6	13.5	158971	2	AP004882	AP004882 Oryza sat
c 26	66.6	13.5	165988	2	AC121212	AC121212 Rattus no
c 27	66.6	13.5	171624	2	AC095621	AC095621 Rattus no
c 28	66.2	13.5	194776	2	AC129706	AC129706 Rattus no
c 29	65.8	13.4	171204	2	AC130936	AC130936 Rattus no
c 30	65.6	13.3	175041	2	AC097133	AC097133 Rattus no
c 31	65.2	13.3	62839	2	AC098999	AC098999 Rattus no
c 32	65.2	13.3	170696	2	AP004658	AP004658 Oryza sat
c 33	65.2	13.3	171204	2	AC130936	AC130936 Rattus no
c 34	65	13.2	184245	2	AC127355	AC127355 Mus muscu
c 35	64.6	13.1	110000	2	AC116301	AC116301 Homo sapi
c 36	64.4	13.1	171852	2	AC058782	AC058782 Homo sapi
c 37	64.4	13.1	181022	2	AC121905	AC121905 Mus muscu
c 38	64.4	13.1	188133	2	AC124248	AC124248 Rattus no
c 39	64.2	13.0	81417	2	AC023265	AC023265 Homo sapi
c 40	64	13.0	69417	2	AC121516	AC121516 Mus muscu
c 41	64	13.0	162183	2	AC097753	AC097753 Rattus no
c 42	63.8	13.0	161163	2	AC115099	AC115099 Homo sapi
c 43	63.8	13.0	178941	2	AC113635	AC113635 Rattus no
c 44	63.6	12.9	129667	2	AC095300	AC095300 Rattus no
c 45	63.6	12.9	135119	2	AC011578	AC011578 Homo sapi

ALIGNMENTS

RESULT 1
AP004313/c

LOCUS
DEFINITION

AP004313
Oryza sativa (japonica cultivar-group) chromosome 7 clone P0440B02,
*** SEQUENCING IN PROGRESS ***, in ordered pieces.

ACCESSION

VERSION
KEYWORDS

SOURCE

ORGANISM

Oryza sativa (japonica cultivar-group)
clone:P0440B02.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.

Pred. No. is the number of results predicted by chance to have a

ORGANISMS

Rattus norvegicus

Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE

REFERENCES

1 (bases 1 to 163381)

(cont'd.)

Murny, D. M., Adams, C., Adio-Oduola, B., Ali-osman, F. R., Allen, C., Alsbrooks, S. L., Amaratunge, H. C., Are, J., Ayale, M., Banks, T., Barbaria, J., Benton, J., Bimage, K., Blankenburg, K., Bonnin, D., Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N. P., Buhay, C., Burch, P., Burkett, C., Burrell, K. L., Byrd, N. C., Carron, T. F., Carter, M., Cavazos, S. R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C., Cleaveland, C. D., Cox, C. C., Coyle, M. D., Dathorne, S. R., David, R., Davilla, M. L., Davis, C., Davy-Carroll, L., Dederich, D. A., Delaney, K. R., Delgado, O., Denn, A. L., Ding, Y., Dinh, H. H., Douthwaite, K. J., Draper, H., Dugan-Rocha, S., Durbin, K. M., Earnhardt, C., Edgar, D., Edwards, C. C., Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J. H., Guevara, M., Gundaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B., Homsi, F., Howard, S., Huber, J., Hulyk, S., Hum, J., Jackson, L. E., Jacobsen, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C., Kratochvic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L. C., Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Loulsegod, H., Lozardo, R. J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhinney, E., McLeod, M. P., Meador, M., Mei, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morqin, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwu, S., Oguh, M., O'Connell, G., O'Connell, N., O'Connell, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L. L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojubakan, I., Royle, M., Ruiz, S., Savery, G., Scher, S., Scott, G., Shen, H., Shoohtari, N., Sisson, I., Sodergren, E., Sonaik, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Taborp, P., Tamerisa, A., Tamerisa, K., Tang, H., Tanton, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usman, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wang, Q., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleciyk, R., Wooden, S., Worley, K., Wu, C., Wu, Y. F., Zhou, J., Zorrilla, S., Nelson, D., Weinstein, G., and Gibbs, R.

TITLE

JOURNAL.

REFERENCE

REFERENCES

AUTHORS
TITLEJOURNAL OF THE AMERICAN MEDICAL ASSOCIATION
PUBLISHED WEEKLY
535 N. Dearborn Ave., Chicago 10, Ill.
Subscription price: \$5.00 per year in advance.
Single copies: 15¢.
Acceptance for mailing at special rate of postage provided for in Section 1103, Act of October 3, 1917. Authorized by Act of October 3, 1917. Postage paid at Chicago, Ill., and at additional mailing offices.
Postmaster: Send address changes in this journal to JOURNAL OF THE AMERICAN MEDICAL ASSOCIATION, 535 N. Dearborn Ave., Chicago 10, Ill.
Second-class postage paid at Chicago, Ill., and at additional mailing offices.
Copyright, 1960, by J. A. M. A. All rights reserved.
Printed at the University of Chicago Press, Chicago, Ill.
Published by the American Medical Association, 535 N. Dearborn Ave., Chicago 10, Ill.
JOURNAL OF THE AMERICAN MEDICAL ASSOCIATION
PUBLISHED WEEKLY
535 N. Dearborn Ave., Chicago 10, Ill.
Subscription price: \$5.00 per year in advance.
Single copies: 15¢.
Acceptance for mailing at special rate of postage provided for in Section 1103, Act of October 3, 1917. Authorized by Act of October 3, 1917. Postage paid at Chicago, Ill., and at additional mailing offices.
Postmaster: Send address changes in this journal to JOURNAL OF THE AMERICAN MEDICAL ASSOCIATION, 535 N. Dearborn Ave., Chicago 10, Ill.
Second-class postage paid at Chicago, Ill., and at additional mailing offices.
Copyright, 1960, by J. A. M. A. All rights reserved.
Printed at the University of Chicago Press, Chicago, Ill.
Published by the American Medical Association, 535 N. Dearborn Ave., Chicago 10, Ill.
JOURNAL OF THE AMERICAN MEDICAL ASSOCIATION
PUBLISHED WEEKLY
535 N. Dearborn Ave., Chicago 10, Ill.
Subscription price: \$5.00 per year in advance.
Single copies: 15¢.
Acceptance for mailing at special rate of postage provided for in Section 1103, Act of October 3, 1917. Authorized by Act of October 3, 1917. Postage paid at Chicago, Ill., and at additional mailing offices.
Postmaster: Send address changes in this journal to JOURNAL OF THE AMERICAN MEDICAL ASSOCIATION, 535 N. Dearborn Ave., Chicago 10, Ill.
Second-class postage paid at Chicago, Ill., and at additional mailing offices.
Copyright, 1960, by J. A. M. A. All rights reserved.
Printed at the University of Chicago Press, Chicago, Ill.
Published by the American Medical Association, 535 N. Dearborn Ave., Chicago 10, Ill.

COMMENT

Direct submission

Unpublished
 Manuscript

2 (bases 1 to 163)

World
J. Sci.

Worley, K.C.
Direct submission

Submitted 19-7-2011

of Molecules and Ions

BI MOLECULAR AND HAN-
Gaulor place HAN-

Baylor Plaza, Hous

Costs: 100,000

Center: Baylor

Center code: H

Web site: <http://www.chem.mcgill.ca/~chem223/>

Contact: hgsc-

-----PROC-----

Center project,

Center clone n

Sum

Sequencing vec

Chemistry: Dye

Assembly progr

Consensus qual

Consensus qual

- * is not known and their order in this sequence record is
- * arbitrary. Gaps between the contigs are represented as
- * runs of N, but the exact sizes of the gaps are unknown.
- * This record will be updated with the finished sequence
- * as soon as it is available and the accession number will
- * be preserved.

1	1027:	contig of 1027 bp in length
1028	1127:	gap of unknown length
1128	2195:	contig of 1068 bp in length
2196	2295:	gap of unknown length
2296	3623:	contig of 1328 bp in length
3624	3723:	gap of unknown length
3724	5144:	contig of 1421 bp in length
5145	5244:	gap of unknown length
5245	6262:	contig of 1018 bp in length
6263	6362:	gap of unknown length
6363	7886:	contig of 1524 bp in length
7887	7986:	gap of unknown length
7987	9581:	contig of 1595 bp in length
9582	9681:	gap of unknown length
9682	11192:	contig of 1511 bp in length
11193	11292:	gap of unknown length
11293	12398:	contig of 1106 bp in length
12399	12498:	gap of unknown length
12499	13696:	contig of 1198 bp in length
13697	13796:	gap of unknown length
13797	14976:	contig of 1180 bp in length
14977	15076:	gap of unknown length
15077	16562:	contig of 1486 bp in length
16563	16662:	gap of unknown length
16663	17680:	contig of 1018 bp in length
17681	17780:	gap of unknown length
17781	18800:	contig of 1020 bp in length
18801	18900:	gap of unknown length
18901	20003:	contig of 1103 bp in length
20004	20103:	gap of unknown length
20104	21681:	contig of 1578 bp in length
21682	21781:	gap of unknown length
21782	22870:	contig of 1089 bp in length
22871	22970:	gap of unknown length
22971	24090:	contig of 1120 bp in length
24091	24190:	gap of unknown length
24191	25426:	contig of 1236 bp in length
25427	25528:	gap of unknown length
25527	26689:	contig of 1163 bp in length
26690	26789:	gap of unknown length
26790	28356:	contig of 1567 bp in length
28357	28456:	gap of unknown length
28457	30062:	contig of 1606 bp in length
30063	30162:	gap of unknown length
30163	31108:	contig of 1444 bp in length
31107	31406:	gap of unknown length
31407	32968:	contig of 1562 bp in length
32969	33068:	gap of unknown length
33069	34291:	contig of 1223 bp in length
34292	34391:	gap of unknown length
34392	35467:	contig of 1076 bp in length
35468	35567:	gap of unknown length
35568	36788:	contig of 1221 bp in length
36789	36888:	gap of unknown length
36889	38584:	contig of 1696 bp in length
38590	38684:	gap of unknown length
38685	40070:	contig of 1386 bp in length
40071	40170:	gap of unknown length
40171	41693:	contig of 1523 bp in length
41694	41793:	gap of unknown length
41794	42956:	contig of 1163 bp in length
42957	43056:	gap of unknown length
43057	44632:	contig of 1596 bp in length
44633	44732:	gap of unknown length
44733	46319:	contig of 1567 bp in length
46320	46419:	gap of unknown length
46420	48471:	contig of 2052 bp in length

http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: W1BR
Web site: http://www-seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu
----- Project Information
Project name: L9871
Center clone name: 124_H_2

* NOTE: This record contains 78 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1 975: contig of 975 bp in length
* 976 1075: gap of 100 bp
* 1076 2100: contig of 1025 bp in length
* 2101 2200: gap of 100 bp
* 2201 3185: contig of 985 bp in length
* 3186 3285: gap of 100 bp
* 3286 4151: contig of 866 bp in length
* 4152 4251: gap of 100 bp
* 4252 5205: contig of 955 bp in length
* 5207 5306: gap of 100 bp
* 5307 6287: contig of 981 bp in length
* 6288 6387: gap of 100 bp
* 6388 7332: contig of 945 bp in length
* 7333 7432: gap of 100 bp
* 7433 8397: contig of 965 bp in length
* 8398 8497: gap of 100 bp
* 8498 9457: contig of 960 bp in length
* 9458 9557: gap of 100 bp
* 9558 10489: contig of 932 bp in length
* 10490 10589: gap of 100 bp
* 10590 11602: contig of 1013 bp in length
* 11603 11702: gap of 100 bp
* 11703 12694: contig of 992 bp in length
* 12695 12794: gap of 100 bp
* 12795 13798: contig of 1004 bp in length
* 13799 13898: gap of 100 bp
* 13899 14866: contig of 968 bp in length
* 14867 14966: gap of 100 bp
* 14967 15930: contig of 964 bp in length
* 15931 16030: gap of 100 bp
* 16031 16981: contig of 951 bp in length
* 16982 17081: gap of 100 bp
* 17082 18043: contig of 962 bp in length
* 18044 18143: gap of 100 bp
* 18144 19119: contig of 976 bp in length
* 19120 19219: gap of 100 bp
* 19220 20173: contig of 954 bp in length
* 20174 20273: gap of 100 bp
* 20274 21232: contig of 959 bp in length
* 21233 21332: gap of 100 bp
* 21333 22227: contig of 895 bp in length
* 22228 22327: gap of 100 bp
* 22328 23303: contig of 976 bp in length
* 23304 23403: gap of 100 bp
* 23404 24393: contig of 990 bp in length
* 24394 24493: gap of 100 bp
* 24494 25495: contig of 1002 bp in length
* 25496 25595: gap of 100 bp
* 25596 26554: contig of 959 bp in length
* 26555 26654: gap of 100 bp
* 26655 27611: contig of 957 bp in length
* 27612 27711: gap of 100 bp

* 27712 28689: contig of 978 bp in length
* 28690 28789: gap of 100 bp
* 28790 29750: contig of 961 bp in length
* 29751 29850: gap of 100 bp
* 29851 30790: contig of 940 bp in length
* 30791 30890: gap of 100 bp
* 30891 31881: contig of 991 bp in length
* 31882 31981: gap of 100 bp
* 31982 32962: contig of 981 bp in length
* 32963 33062: gap of 100 bp
* 33063 34046: contig of 984 bp in length
* 34047 34146: gap of 100 bp
* 34147 35108: contig of 962 bp in length
* 35109 35208: gap of 100 bp
* 35209 36136: contig of 928 bp in length
* 36137 36236: gap of 100 bp
* 36237 37202: contig of 966 bp in length
* 37203 37302: gap of 100 bp
* 37303 38272: contig of 970 bp in length
* 38273 38372: gap of 100 bp
* 38373 39319: contig of 947 bp in length
* 39320 39419: gap of 100 bp
* 39420 40376: contig of 957 bp in length
* 40377 40476: gap of 100 bp
* 40477 41464: contig of 988 bp in length
* 41465 41564: gap of 100 bp
* 41565 42492: contig of 928 bp in length
* 42493 42592: gap of 100 bp
* 42593 43578: contig of 986 bp in length
* 43579 43678: gap of 100 bp
* 43679 44686: contig of 1008 bp in length
* 44687 44786: gap of 100 bp
* 44787 45745: contig of 959 bp in length
* 45746 45845: gap of 100 bp
* 45846 46827: contig of 982 bp in length
* 46828 46927: gap of 100 bp
* 46928 47910: contig of 983 bp in length
* 47911 48010: gap of 100 bp
* 48011 48969: contig of 959 bp in length
* 48970 49065: gap of 100 bp
* 49070 50028: contig of 959 bp in length
* 50029 50128: gap of 100 bp
* 50129 51053: contig of 925 bp in length
* 51054 51153: gap of 100 bp
* 51154 52111: contig of 958 bp in length
* 52112 52211: gap of 100 bp
* 52212 53189: contig of 978 bp in length
* 53190 54289: gap of 100 bp
* 54290 54271: contig of 982 bp in length
* 54272 54371: gap of 100 bp
* 54372 55346: contig of 965 bp in length
* 55337 55436: gap of 100 bp
* 55437 56366: contig of 930 bp in length
* 56367 56466: gap of 100 bp
* 56467 57415: contig of 949 bp in length
* 57416 57515: gap of 100 bp
* 57516 58501: contig of 986 bp in length
* 58502 58601: gap of 100 bp
* 58602 59567: contig of 966 bp in length
* 59568 59667: gap of 100 bp
* 59668 60585: contig of 918 bp in length
* 60586 60685: gap of 100 bp
* 60686 61646: contig of 961 bp in length
* 61647 61746: gap of 100 bp
* 61747 62739: contig of 993 bp in length
* 62740 62839: gap of 100 bp
* 62840 63824: contig of 985 bp in length
* 63825 63924: gap of 100 bp
* 63925 64903: contig of 979 bp in length
* 64904 65003: gap of 100 bp
* 65004 65963: contig of 960 bp in length
* 65964 66063: gap of 100 bp
* 66064 67010: contig of 947 bp in length

* 67011 67110: gap of 100 bp	
* 67111 68077: contig of 967 bp in length	
* 68078 68177: gap of 100 bp	
* 68178 69147: contig of 970 bp in length	
* 69148 69247: gap of 100 bp	
* 69248 70258: contig of 1011 bp in length	
* 70259 70358: gap of 100 bp	
* 70359 71286: contig of 928 bp in length	
* 71287 71386: gap of 100 bp	
* 71387 72279: contig of 893 bp in length	
* 72280 72379: gap of 100 bp	
* 72380 73317: contig of 938 bp in length	
* 73318 73417: gap of 100 bp	
Query Match 15.1%; Score 74.4; DB 2; Length 82815;	
Best Local Similarity 48.9%; Pred. No. 0.00036;	
Matches 159; Conservative 0; Mismatches 166; Indels 0; Gaps 0;	
QY 17 CTCCGCGCGACATTAAACCTTGCTCCGACAGCGCGCGACTCTCCCAATGCTC 76	
Db 66610 CC 66669	
QY 77 TCCTCTTCCTCCACCCACTCGGCGACCTCGGCCCTTCCCGCGGCTCGGCTCCGCAAAAC 136	
Db 66670 CC 66729	
QY 137 CTTAACCCCGCTCTCGCTCTCCGCTCTCGGCGCCCTCTCGGCCGCTGCTCGCGGTC 196	
Db 66730 CC 66789	
QY 197 TTGAGCGCGAGTGGGCGCCAGCCGATGAGTACGTTAGGCGCTCTGTTGCGCGTGGGGCG 256	
Db 66790 CGCGGGCGGGCG 66849	
QY 257 GCGGCGGGAGAGCTGGGCGAGTGGGGAGTTGGCGACGAGCGGCTGGTGGAGTGGCG 316	
Db 66850 GGGGCGGGGGCGCGCGCGCGCGCGGGGNNCGCCCGCGCTCGGGCGGGGGGGGGG 66909	
QY 317 CAACGGCGCGGACGCTGCTGGGG 341	
Db 66910 CGG 66934	
RESULT 7	
AC121439	
LOCUS	Rattus norvegicus clone CH230-247E4, *** SEQUENCING IN PROGRESS
DEFINITION	***, 56 unordered pieces.
ACCESSION	AC121439
VERSION	AC121439.2 GI:21321741
KEYWORDS	HTG; HTGS_PHASE1.
SOURCE	Rattus norvegicus.
ORGANISM	Rattus norvegicus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
AUTHORS	1 (bases 1 to 158897)
	Munzy D.M., Adams C., Adio-Oduola B., Ali-osman, F.R., Allen C., Alsbrooks S.L., Amarantunga H.C., Are J.R., Ayelle M., Banks T., Barbara J., Benton J., Bimaga K., Blankenburg K., Bonnin D., Bouck J., Bowie S., Brivea M., Brown E., Brown M., Bryant N.P., Buhaq C., Burch P., Burkett C., Burrell K.L., Byrd N.C., Caron T.F., Carter M., Cavazos S.R., Chacko J., Chavez D., Chen G., Chen R., Chen Z., Chowdhry I., Christopoulos C., Cleveland C.D., Cox C., Coyle M.D., Dathorne S.R., David R., Davila M.L., Davis C., Davy-Carroll L., Dederich D.A., Delaney K.R., Delgado O., Denn A.L., Ding Y., Dinoh H.H., Douthwaite K.J., Draper H., Dugan-Rocha S., Durbin K.J., Earnhart C., Edgar D., Edwards C.C., Elhaj C., Escotto M., Falls T., Ferraguto D., Flagg N., Ford J., Foster P., Frantz P., Gabisi A., Gao J., Garcia A., Garner T., Garza N., Gill R., Gorrell J.H., Guevara W., Gunaratne P., Hale S., Hamilton K., Harris C., Harris K., Hart M., Havlak P., Hayes A., Hernandez J.,

Hernandez, O., Hodgson, A., Hogue, M., Holloway, C., Hollins, B., Honsi, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Loulseged, H., Lozado, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhiney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S., Ogih, M., Okwuonu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojokan, I., Rolfe, M., Ruiz, S., Saverly, G., Scheraga, S., Scott, G., Shen, H., Shoohtari, N., Sisson, I., Sodergren, E., Sonaika, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wang, Q., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczek, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D., Weinstein, G., and Gibbs, R.

Unpublished
2 (bases 1 to 158897)
Worley, K.C.
Direct Submission
Submitted (18-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 158897)
Worley, K.C.
Direct Submission
Submitted (24-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Jun 4, 2002 this sequence version replaced gi:20976389.

Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu

Project Information
Center project name: GXXP
Center clone name: CH230-247E4

Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 125365 bases at least Q40
Consensus quality: 129911 bases at least Q30
Consensus quality: 132859 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank/draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently consists of 56 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.
* 1137: contig of 1137 bp in length
* 1237: gap of unknown length
* 1238: contig of 1153 bp in length
* 2390: gap of unknown length
* 2491: contig of 1093 bp in length
* 3583: gap of unknown length
* 3684: contig of 1137 bp in length
* 4821: gap of unknown length

TITLE
JOURNAL
REFERENCE
AUTHORS

Mus musculus, clone RP23-467G14
Unpublished
2 (bases 1 to 68330)
Birken, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L.,
Boukhalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J.,
Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A.,
Cook, A., Cooke, P., DeArellano, K., Dewar, K., Diaz, J. S., Dodge, S.,
Faro, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S.,
Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
Kamat, A., Karatas, A., Kells, C., LaRoque, K., Lamazares, R.,
Landers, T., Lehoczy, J., Levine, R., Lindblad-Toh, K., Liu, G.,
MacLean, C., MacDonald, P., Major, J., Marquis, N., Matthews, C.,
McCarthy, M., McEwan, P., McKernan, K., Meldrum, J., Meneus, L.,
Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R.,
Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D.,
Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,
Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,
Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupbach, R.,
Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
Topham, K., Travers, M., Travis, N., Trigglio, J., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (25-MAR-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

TITLE
JOURNAL
COMMENT

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L22810
Center clone name: 467_G_14

* NOTE: This record contains 85 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1 647: contig of 647 bp in length
648 747: gap of 100 bp
748 1422: contig of 675 bp in length
1423 1522: gap of 100 bp
1523 2229: contig of 707 bp in length
2230 2329: gap of 100 bp
2330 3001: contig of 672 bp in length
3002 3101: gap of 100 bp
3102 3783: contig of 682 bp in length
3784 3883: gap of 100 bp in length
3884 4586: contig of 703 bp in length
4587 4686: gap of 100 bp
4687 5381: contig of 695 bp in length
5382 5481: gap of 100 bp
5482 6190: contig of 709 bp in length
6191 6290: gap of 100 bp
6291 7003: contig of 713 bp in length
7004 7103: gap of 100 bp
7104 7824: contig of 721 bp in length
7825 7924: gap of 100 bp
7925 8662: contig of 738 bp in length
8663 8762: gap of 100 bp

8763 9402: contig of 640 bp in length
9403 9502: gap of 100 bp
9503 10235: contig of 733 bp in length
10236 10333: gap of 100 bp
10336 11055: contig of 720 bp in length
11056 11155: gap of 100 bp
11156 11861: contig of 706 bp in length
11862 12653: contig of 692 bp in length
12654 12753: gap of 100 bp
12754 13445: contig of 692 bp in length
13446 14335: contig of 690 bp in length
14336 15046: contig of 711 bp in length
15047 15146: gap of 100 bp
15147 15853: contig of 707 bp in length
15854 1597: gap of 100 bp
1598 1632: contig of 679 bp in length
1633 16732: gap of 100 bp
16733 17441: contig of 709 bp in length
17442 17541: gap of 100 bp
17542 18270: contig of 729 bp in length
18271 18370: gap of 100 bp
18371 19048: contig of 678 bp in length
19049 19148: gap of 100 bp
19149 19851: contig of 703 bp in length
19852 19951: gap of 100 bp
19952 20661: contig of 710 bp in length
20662 20761: gap of 100 bp
20762 21473: contig of 712 bp in length
21474 21573: gap of 100 bp
21574 22286: contig of 713 bp in length
22287 23086: contig of 700 bp in length
23087 23186: gap of 100 bp
23187 23920: contig of 734 bp in length
23921 24020: gap of 100 bp
24021 24713: contig of 693 bp in length
24714 24813: gap of 100 bp
24814 25540: contig of 727 bp in length
25541 25640: gap of 100 bp
25641 26364: contig of 724 bp in length
26365 26464: gap of 100 bp
26465 27180: contig of 716 bp in length
27181 27280: gap of 100 bp
27281 27977: contig of 697 bp in length
27978 28077: gap of 100 bp
28078 28769: contig of 692 bp in length
28770 28866: gap of 100 bp
28870 29559: contig of 690 bp in length
29560 29659: gap of 100 bp
29660 30357: contig of 698 bp in length
30358 30457: gap of 100 bp
30458 31165: contig of 708 bp in length
31166 31265: gap of 100 bp
31266 31948: contig of 683 bp in length
31949 32048: gap of 100 bp
32049 32768: contig of 720 bp in length
32769 32868: gap of 100 bp
32869 33548: contig of 680 bp in length
33549 33648: gap of 100 bp
33649 34357: contig of 709 bp in length
34358 34457: gap of 100 bp
34458 35187: contig of 730 bp in length
35188 35287: gap of 100 bp
35288 35990: contig of 703 bp in length
35991 36090: gap of 100 bp
36091 36767: contig of 677 bp in length
36768 36867: gap of 100 bp
36868 37575: contig of 708 bp in length
37576 37675: gap of 100 bp
37676 38379: contig of 704 bp in length

--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

RESULT 9
AC126076
LOCUS
DEFINITION Rattus norvegicus clone CH230-143P6, *** SEQUENCING IN PROGRESS


```
OY 249 CTGGGGCGCGCGGAGAGCTCGCGAGTGGGGAGTTGGCGACGAGCGGCTGGTG 308
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 48989 GCGGGGGGGGGCGGGGGCGCGGGCGGGGGGGGGGGGGGGGGGGGGGG 49048
OY 309 AGGTGGCGCAACGGCGGGGAGCTGTCTGGGAGGTGTC 349
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 49049 CGGGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGCGGCGCTC 49089

RESULT 10
AC027353/C
LOCUS AC027353 101509 bp DNA linear HTG 26-SEP-2000
DEFINITION Homo sapiens chromosome 16 clone RP11-167D21, WORKING DRAFT
AC027353
VERSION AC027353.4 GI:10305180
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 101509)
DOE Joint Genome Institute.
Sequencing of Human Chromosome 16
Unpublished
2 (bases 1 to 101509)
DOE Joint Genome Institute.
Direct Submission
Submitted (30-MAR-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Sep 26, 2000 this sequence version replaced g: 9090641.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 0
Center clone name: RPCI-11_167D21
-----
Summary Statistics
Consensus quality: 19476 bases at least Q40
Consensus quality: 22258 bases at least Q30
Consensus quality: 25343 bases at least Q20
Estimated insert size: 26000; agarose-fp estimation
Quality coverage: 15.03 in Q20 bases; agarose-fp estimation
Quality coverage: 4.12 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 68 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1470: contig of 1470 bp in length
* 1471 1570: gap of unknown length
* 1571 2728: contig of 1158 bp in length
* 2729 2828: gap of unknown length
* 2829 3955: contig of 1127 bp in length
* 3956 4055: gap of unknown length
* 4056 5273: contig of 1218 bp in length
* 5274 5374: gap of unknown length
* 5374 6455: contig of 1082 bp in length
* 6456 6555: gap of unknown length
* 6556 7722: contig of 1167 bp in length
* 7723 7823: gap of unknown length
* 7823 8994: contig of 1172 bp in length
* 8995 9095: gap of unknown length
* 9095 10258: contig of 1164 bp in length
* 10259 10359: gap of unknown length
* 10359 11529: contig of 1171 bp in length

11530
11629: gap of unknown length
11630 12722: contig of 1143 bp in length
12723 12872: gap of unknown length
12873 13954: contig of 1082 bp in length
13955 14054: gap of unknown length
14055 15127: contig of 1073 bp in length
15128 15227: gap of unknown length
15228 16411: contig of 1184 bp in length
16412 16511: gap of unknown length
16512 17557: contig of 1046 bp in length
17558 17657: gap of unknown length
17658 18766: contig of 1109 bp in length
18767 18866: gap of unknown length
18867 20034: contig of 1168 bp in length
20035 20134: gap of unknown length
20135 21336: contig of 1202 bp in length
21337 21436: gap of unknown length
21437 22545: contig of 1109 bp in length
22546 22645: gap of unknown length
22646 23755: contig of 1110 bp in length
23756 23855: gap of unknown length
23856 24927: contig of 1072 bp in length
24928 25027: gap of unknown length
25028 26153: contig of 1126 bp in length
26154 26253: gap of unknown length
26254 27455: contig of 1202 bp in length
27456 27555: gap of unknown length
27556 28679: contig of 1124 bp in length
28680 28779: gap of unknown length
28780 29893: contig of 1114 bp in length
29894 31172: contig of 1179 bp in length
31173 31272: gap of unknown length
31273 32388: contig of 1116 bp in length
32389 32488: gap of unknown length
32489 33651: contig of 1162 bp in length
33652 33750: gap of unknown length
33751 34987: contig of 1237 bp in length
34988 35087: gap of unknown length
35088 36153: contig of 1066 bp in length
36154 36253: gap of unknown length
36254 37421: contig of 1168 bp in length
37422 37521: gap of unknown length
37522 38177: contig of 1196 bp in length
38178 38817: gap of unknown length
38818 39915: contig of 1098 bp in length
39916 40015: gap of unknown length
40016 41215: contig of 1200 bp in length
41216 41915: gap of unknown length
41916 42440: contig of 1035 bp in length
42441 42450: gap of unknown length
42451 43607: contig of 1157 bp in length
43608 43707: gap of unknown length
43708 44880: contig of 1173 bp in length
44881 44980: gap of unknown length
44981 46095: contig of 1115 bp in length
46096 46195: gap of unknown length
46196 47463: contig of 1268 bp in length
47464 47563: gap of unknown length
47564 48771: contig of 1208 bp in length
48772 48871: gap of unknown length
48872 50000: contig of 1129 bp in length
50001 50100: gap of unknown length
50101 51241: contig of 1141 bp in length
51242 51341: gap of unknown length
51342 52439: contig of 1098 bp in length
52440 52539: gap of unknown length
52540 53719: contig of 1180 bp in length
53720 53819: gap of unknown length
53820 54972: contig of 1153 bp in length
54973 55072: gap of unknown length
55073 56256: contig of 1184 bp in length
56257 56356: gap of unknown length
```

[illegible]

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

3 (bases 1 to 151358)
Worley, K.C.
Direct Submission
Submitted (18-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GVRX
Center clone name: CH230-77F8
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 92473 bases at least Q40
Consensus quality: 98495 bases at least Q30
Consensus quality: 102505 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_drift_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 63 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 1032: contig of 1032 bp in length
* 1033 1132: gap of unknown length
* 1133 2134: contig of 1002 bp in length
* 2135 2234: gap of unknown length
* 2235 3255: contig of 1021 bp in length
* 3256 3355: gap of unknown length
* 3356 4398: contig of 1043 bp in length
* 4399 4498: gap of unknown length
* 4499 5857: contig of 1359 bp in length
* 5858 5957: gap of unknown length
* 5958 7206: contig of 1249 bp in length
* 7207 7306: gap of unknown length
* 7307 8379: contig of 1073 bp in length
* 8380 8479: gap of unknown length
* 8480 9982: contig of 1503 bp in length
* 9983 10082: gap of unknown length
* 10083 11373: contig of 1291 bp in length
* 11374 11473: gap of unknown length
* 11474 12776: contig of 1303 bp in length
* 12777 12876: gap of unknown length
* 12877 14025: contig of 1149 bp in length
* 14026 14125: gap of unknown length
* 14126 15578: contig of 1453 bp in length
* 15579 15678: gap of unknown length
* 15679 16957: contig of 1279 bp in length
* 16958 17057: gap of unknown length
* 17058 18241: contig of 1184 bp in length
* 18242 18341: gap of unknown length
* 18342 19373: contig of 1032 bp in length
* 19374 19473: gap of unknown length
* 19474 21050: contig of 1577 bp in length
* 21051 21150: gap of unknown length
* 21151 22862: contig of 1712 bp in length
* 22863 24563: contig of 1601 bp in length
* 24564 24663: gap of unknown length
* 24664 25683: contig of 1020 bp in length
* 25684 25783: gap of unknown length
* 25784 26831: contig of 1048 bp in length
* 26832 26931: gap of unknown length
* 26932 28243: contig of 1312 bp in length

28244 28343: gap of unknown length
28344 28758: contig of 1415 bp in length
28759 29759: gap of unknown length
29859 31174: contig of 1316 bp in length
31175 31274: gap of unknown length
31275 33433: contig of 2159 bp in length
33434 33533: gap of unknown length
33534 35127: contig of 1594 bp in length
35128 35227: gap of unknown length
35228 36393: contig of 1166 bp in length
36394 36494: gap of unknown length
36494 38614: contig of 2121 bp in length
38615 38714: gap of unknown length
38715 40665: contig of 1951 bp in length
40666 40765: gap of unknown length
40766 41885: contig of 1120 bp in length
41886 44086: contig of 2101 bp in length
44087 44186: gap of unknown length
44187 46007: contig of 1821 bp in length
46008 46107: gap of unknown length
46108 47865: contig of 1758 bp in length
47866 47965: gap of unknown length
47966 50121: contig of 2056 bp in length
50122 50121: gap of unknown length
50122 51983: contig of 1862 bp in length
51984 52083: gap of unknown length
52084 54300: contig of 2217 bp in length
54301 54400: gap of unknown length
54401 55910: contig of 1510 bp in length
55911 56010: gap of unknown length
56011 57542: contig of 1532 bp in length
57543 57642: gap of unknown length
57643 59295: contig of 1653 bp in length
59296 59395: gap of unknown length
59396 62493: contig of 2998 bp in length
62494 64229: contig of 1736 bp in length
64230 64329: gap of unknown length
64330 66089: contig of 1760 bp in length
66090 66189: gap of unknown length
66190 67309: contig of 2720 bp in length
67309 68910: gap of unknown length
68910 71491: contig of 2482 bp in length
71492 71591: gap of unknown length
71592 73355: contig of 1764 bp in length
73356 73455: gap of unknown length
73456 75777: contig of 2322 bp in length
75778 75877: gap of unknown length
75878 78492: contig of 2515 bp in length
78493 81849: contig of 3357 bp in length
81850 81949: gap of unknown length
81950 84203: contig of 2253 bp in length
84203 84303: gap of unknown length
84303 86896: contig of 2594 bp in length
86897 86996: gap of unknown length
86997 90467: contig of 3471 bp in length
90468 90567: gap of unknown length
90568 93764: contig of 3197 bp in length
93765 93864: gap of unknown length
93865 96842: contig of 2978 bp in length
96843 96942: gap of unknown length
96943 99989: contig of 3047 bp in length
99990 100089: gap of unknown length
100090 103155: contig of 3066 bp in length
103156 103255: gap of unknown length

Query Match 14.2%; Score 69.8; DB 2; Length 151358;
Best Local Similarity 51.9%; Pred. No. 0.0025;
Matches 149; Conservative 0; Mismatches 138; Indels 0; Gaps 0;

Oy 84 CCTGCTCCACCCACTCGGCCACCTCGGCCTTCCCGGGCTCGCCTCCGCAACCCCTAACCC 143

Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,
 Nguyen, N., Nickerson, E., Nwokenkwo, S., Oguh, M., Okwuonu, G.,
 Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,
 Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y.,
 Rivas, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savery, G.,
 Scherer, S., Scott, G., Shen, H., Shoostari, N., Sisson, I.,
 Sodergren, E., Sonaik, T., Sparks, A., Stanley, H., Stone, H.,
 Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,
 Tansley, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,
 Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wang, Q.,
 Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,
 Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K.,
 Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
 Weinstein, G., and Gibbs, R.
 Direct Submission
 Unpublished
 2 (bases 1 to 174031)
 Worley, K.C.
 Direct Submission
 Submitted (15-APR-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 174031)
 Worley, K.C.
 Direct Submission
 Submitted (18-JUL-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Jul 14, 2002 this sequence version replaced gi:20330629.
 ----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GUGG
 Center clone name: CH230-173H17
 ----- Summary Statistics
 Sequencing vector: plasmid;
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 107398 bases at least Q40
 Consensus quality: 119658 bases at least Q30
 Consensus quality: 127269 bases at least Q20

 * NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 58 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 *
 * 1
 * 1025 1024: contig of 1024 bp in length
 * 1125 1124: gap of unknown length
 * 2168 2167: contig of 1043 bp in length
 * 2268 2267: gap of unknown length
 * 3415 3414: contig of 1147 bp in length
 * 3515 3514: gap of unknown length
 * 4715 4714: contig of 1200 bp in length
 * 4815 4814: gap of unknown length
 * 6354 6353: contig of 1539 bp in length
 * 6454 6453: gap of unknown length
 * 7916 7915: contig of 1462 bp in length
 * 8016 8015: gap of unknown length
 * 9360 9359: contig of 1344 bp in length
 * 9460 9459: gap of unknown length
 * 11083 11082: contig of 1623 bp in length
 * 11183 11182: gap of unknown length
 * 12370 12369: contig of 1187 bp in length
 * 12469 12469: gap of unknown length

12470 13330: contig of 1461 bp in length
 13931 14030: gap of unknown length
 14031 15253: contig of 1223 bp in length
 15254 15353: gap of unknown length
 15354 16696: contig of 1343 bp in length
 16697 16796: gap of unknown length
 16797 18403: contig of 1607 bp in length
 18404 18503: gap of unknown length
 18504 19776: contig of 1273 bp in length
 19777 19876: gap of unknown length
 19877 21648: contig of 1772 bp in length
 21649 21748: gap of unknown length
 21749 23094: contig of 1346 bp in length
 23095 23194: gap of unknown length
 23195 25092: contig of 1898 bp in length
 25093 25192: gap of unknown length
 25193 26802: contig of 1610 bp in length
 26803 26902: gap of unknown length
 26903 28726: contig of 1824 bp in length
 28727 28826: gap of unknown length
 28828 31064: contig of 2238 bp in length
 31065 33254: contig of 2090 bp in length
 33255 33354: gap of unknown length
 33355 35763: contig of 2409 bp in length
 35764 35863: gap of unknown length
 35864 36989: contig of 1126 bp in length
 36990 37089: gap of unknown length
 37090 38703: contig of 1614 bp in length
 38704 38803: gap of unknown length
 38804 40934: contig of 2131 bp in length
 40935 41034: gap of unknown length
 41035 42978: contig of 1944 bp in length
 42979 45214: contig of 2136 bp in length
 43079 45215: gap of unknown length
 45215 45314: gap of unknown length
 45315 47780: contig of 2466 bp in length
 47781 47880: gap of unknown length
 47881 50209: contig of 2329 bp in length
 50210 50309: gap of unknown length
 50310 52364: contig of 2055 bp in length
 52365 52464: gap of unknown length
 52465 54739: contig of 2275 bp in length
 54740 54839: gap of unknown length
 54840 56920: contig of 2081 bp in length
 56921 57020: gap of unknown length
 57021 59676: contig of 2656 bp in length
 59677 59776: gap of unknown length
 59777 62058: gap of unknown length
 62059 64481: contig of 2323 bp in length
 62159 64481: contig of 2323 bp in length
 64482 64581: gap of unknown length
 64582 67737: contig of 3156 bp in length
 67738 67837: gap of unknown length
 67838 69856: contig of 2019 bp in length
 69857 69956: gap of unknown length
 69957 72698: contig of 2742 bp in length
 72699 72798: gap of unknown length
 72799 75901: contig of 3102 bp in length
 75901 76000: gap of unknown length
 76001 78614: contig of 2614 bp in length
 78615 78714: gap of unknown length
 78715 81699: contig of 2985 bp in length
 81699 81700: gap of unknown length
 81700 85909: contig of 4110 bp in length
 85910 86009: gap of unknown length
 86010 89896: contig of 3787 bp in length
 89897 93635: contig of 3739 bp in length
 93636 93735: gap of unknown length
 93736 97658: contig of 3923 bp in length
 97659 97758: gap of unknown length
 97759 101045: contig of 3287 bp in length

* 4069 4168: gap of unknown length
* 4169 5608: contig of 1440 bp in length
* 5609 7127: gap of unknown length
* 7128 7227: contig of 1419 bp in length
* 7228 8577: gap of unknown length
* 8578 8677: contig of 1350 bp in length
* 8678 9793: gap of unknown length
* 9794 9893: contig of 1116 bp in length
* 9894 10907: gap of unknown length
* 10908 11007: contig of 1014 bp in length
* 11008 12349: gap of unknown length
* 12350 12449: contig of 1342 bp in length
* 12450 13548: gap of unknown length
* 13549 13649: contig of 1099 bp in length
* 13650 14731: gap of unknown length
* 14732 14830: contig of 1082 bp in length
* 14831 16231: gap of unknown length
* 16232 16351: contig of 1421 bp in length
* 16352 18478: gap of unknown length
* 18479 20188: contig of 2127 bp in length
* 20189 20288: gap of unknown length
* 20289 21551: contig of 1610 bp in length
* 21552 21651: gap of unknown length
* 21653 23305: contig of 1263 bp in length
* 23306 23405: gap of unknown length
* 23406 25560: contig of 1654 bp in length
* 25561 25660: gap of unknown length
* 25661 27283: contig of 2155 bp in length
* 27284 27384: gap of unknown length
* 27385 29246: contig of 1623 bp in length
* 29247 29346: gap of unknown length
* 29347 31576: contig of 1863 bp in length
* 31577 33133: gap of unknown length
* 33134 33413: contig of 2230 bp in length
* 33414 35876: contig of 1637 bp in length
* 35877 35976: gap of unknown length
* 35977 37553: contig of 2463 bp in length
* 37554 37653: gap of unknown length
* 37654 39176: contig of 1577 bp in length
* 39177 39276: gap of unknown length
* 39277 40741: contig of 1523 bp in length
* 40742 40841: contig of 1465 bp in length
* 40842 42390: gap of unknown length
* 42391 42491: contig of 1549 bp in length
* 42492 44654: gap of unknown length
* 44655 44754: contig of 2164 bp in length
* 44755 46171: gap of unknown length
* 46172 46271: contig of 1417 bp in length
* 46272 48302: gap of unknown length
* 48303 48403: contig of 2031 bp in length
* 48404 50074: gap of unknown length
* 50075 50174: contig of 1672 bp in length
* 50175 52341: gap of unknown length
* 52342 52441: contig of 2167 bp in length
* 52442 54601: gap of unknown length
* 54602 54701: contig of 2160 bp in length
* 54702 57006: gap of unknown length
* 57007 57106: contig of 2305 bp in length
* 57107 59815: contig of 2609 bp in length
* 59816 62344: gap of unknown length
* 62345 62445: contig of 2529 bp in length
* 62446 66001: gap of unknown length
* 66002 66101: contig of 3557 bp in length
* 66102 68787: contig of 2686 bp in length.

FEATURES
source

Location/Qualifiers
1. .68787
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/clone="CH230-137B1"

BASE COUNT 18229 a 14391 c 13550 g 17475 t 5142 others
ORIGIN
Query Match 13.7%; Score 67.4; DB 2; Length 68787;
Best Local Similarity 54.1%; Pred. No. 0.0071;
Matches 157; Conservative 0; Mismatches 132; Indels 1; Gaps 1;
QY 16 TCCTCCGCGCGACATTAAACCTTGTCTCCGACAAAGCGCGGCTCTCTCCCAATGCT 75
Db 31191 TCCTCNCCTCCGCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 31250
QY 76 CT 135
Db 31251 CCCCCCTCCG 31310
QY 136 CCTTAACCCCGCT 195
Db 31311 CCCCCCG 31370
QY 196 CTTGAGCGGAGTGGGCG 255
Db 31371 GCGGTGCG 31429
QY 256 GCGCGCGCGGAGAGTCTCGGCGAGTGGGGAGTGGGGAGTGGGGAGTGGGGAGTGGGG 305
Db 31430 CG 31479

RESULT 15
LOCUS AC126730
DEFINITION Rattus norvegicus clone CH230-6H4, *** SEQUENCING IN PROGRESS ***,
69 unordered pieces.
AC126730
AC126730.2 GI:21886838
HTG: HTGS_PHASE1.
KEYWORDS Rattus norvegicus.
SOURCE Rattus norvegicus.
ORGANISM Rattus

REFERENCE 1 (bases 1 to 130899)
AUTHORS Muzny, D.M., Adams, C., Ali-ouman, F.R., Allen, C., Albrooks, S.L., Amaral, H.C., Are, J.R., Ayale, M., Banks, T., Barbra, J., Benton, J., Bimane, K., Blankenburg, K., Bonin, D., Bouck, J., Bowie, S., Brieva, M., Brown, M., Brown, N.P., Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, K., Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A., Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hogue, M., Holloway, C., Hollins, B., Homs, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvan, J., Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Louised, H., Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhiney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M., Miner, G., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, Z., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenwo, S., Ogih, M., Okwuonu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojebokan, I., Rolfe, M., Ruiz, S., Savary, G.,

Scherer, S., Scott, G., Shen, H., Shooshtari, N., Sisson, I.,
 Sodergren, E., Sonaika, T., Sparks, A., Stanley, H., Stone, H.,
 Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,
 Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,
 Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wang, Q.,
 Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,
 Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K.,
 Wu, C., Wu, Y., Wu, Y. F., Zhou, J., Zorrilla, S., Nelson, D.,
 Weinstein, G. and Gibbs, R.
 Direct Submission
 Unpublished
 2 (bases 1 to 130899)
 Worley, K.C.
 Direct Submission
 Submitted (09-JUL-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 130899)
 Worley, K.C.
 Direct Submission
 Submitted (23-JUL-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Jul 17, 2002 this sequence version replaced gi:21716710.

 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu

 Project Information
 Center project name: GBWP
 Center Clone name: CH230-6H4

 Summary Statistics
 Sequencing vector: Plasmid;
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 71039 bases at least Q40
 Consensus quality: 77045 bases at least Q30
 Consensus quality: 80907 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
 (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 69 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1
 1028: contig of 1027 bp in length
 1127: gap of unknown length
 1128: contig of 1145 bp in length
 2271: contig of 1145 bp in length
 2272: gap of unknown length
 2373: contig of 1026 bp in length
 3399: gap of unknown length
 3499: contig of 1304 bp in length
 4802: gap of unknown length
 4803: contig of 1091 bp in length
 5993: gap of unknown length
 6094: contig of 1151 bp in length
 7245: gap of unknown length
 7345: contig of 1536 bp in length
 8881: gap of unknown length
 8981: contig of 1007 bp in length
 9988: gap of unknown length
 10088: contig of 1255 bp in length
 11343: gap of unknown length
 11443: contig of 1118 bp in length
 12561: gap of unknown length
 12661: contig of 1224 bp in length
 13885: gap of unknown length
 15064: contig of 1080 bp in length

15065: gap of unknown length
 15164: contig of 1382 bp in length
 16347: gap of unknown length
 16348: contig of 1867 bp in length
 18513: gap of unknown length
 18514: contig of 1092 bp in length
 19705: gap of unknown length
 19706: contig of 1465 bp in length
 19806: gap of unknown length
 21271: contig of 1230 bp in length
 21371: gap of unknown length
 22601: contig of 1041 bp in length
 22701: gap of unknown length
 23742: contig of 1104 bp in length
 23842: gap of unknown length
 24946: contig of 2213 bp in length
 25046: gap of unknown length
 27259: contig of 1259 bp in length
 27359: gap of unknown length
 28618: contig of 1175 bp in length
 28718: gap of unknown length
 29893: contig of 1404 bp in length
 29993: gap of unknown length
 31397: contig of 1008 bp in length
 31496: gap of unknown length
 32505: contig of 1396 bp in length
 32605: gap of unknown length
 34001: contig of 1757 bp in length
 34101: gap of unknown length
 35858: contig of 1420 bp in length
 35958: gap of unknown length
 37378: contig of 1020 bp in length
 37478: gap of unknown length
 38498: contig of 1096 bp in length
 38598: gap of unknown length
 39694: contig of 2709 bp in length
 39794: gap of unknown length
 42503: contig of 1821 bp in length
 42603: gap of unknown length
 44224: contig of 1707 bp in length
 44244: gap of unknown length
 46231: contig of 1429 bp in length
 46331: gap of unknown length
 47759: contig of 1349 bp in length
 47860: gap of unknown length
 49209: contig of 1052 bp in length
 49309: gap of unknown length
 50361: contig of 1797 bp in length
 50461: gap of unknown length
 52257: contig of 1790 bp in length
 52258: gap of unknown length
 52358: contig of 1155 bp in length
 54148: gap of unknown length
 54248: contig of 1506 bp in length
 55403: gap of unknown length
 55503: contig of 1439 bp in length
 57109: gap of unknown length
 57109: contig of 1512 bp in length
 58548: gap of unknown length
 58648: contig of 1469 bp in length
 60160: gap of unknown length
 60260: contig of 2557 bp in length
 61729: gap of unknown length
 61829: contig of 1601 bp in length
 64396: gap of unknown length
 64996: contig of 1516 bp in length
 66097: gap of unknown length
 67113: contig of 1185 bp in length
 67813: gap of unknown length
 68998: contig of 1770 bp in length
 69098: gap of unknown length
 70968: contig of 1162 bp in length
 72130: gap of unknown length
 72229: contig of 1162 bp in length

GenCore version 5.1.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 17, 2002, 14:17:57 : Search time 2693 Seconds
(without alignments)
5316.963 Million cell updates/sec

Title: US-09-686-522c-13

Perfect score: 492

Sequence: 1 agctccgagcgtcattctcc.....agccaatttggngaagaga 492

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*

2: gb_htg.*

3: gb_in.*

4: gb_om.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_sts.*

12: gb_sy.*

13: gb_un.*

14: gb_vl.*

15: em_ba.*

16: em_fun.*

17: em_hum.*

18: em_in.*

19: em_mu.*

20: em_om.*

21: em_or.*

22: em_ov.*

23: em_pat.*

24: em_ph.*

25: em_pl.*

26: em_ro.*

27: em_sts.*

28: em_un.*

29: em_vl.*

30: em_htg_hum.*

31: em_htg_inv.*

32: em_htg_other.*

33: em_htg_mus.*

34: em_htg_pln.*

35: em_htg_rdd.*

36: em_htg_mam.*

37: em_htg_vrt.*

38: em_sy.*

39: em_htgo_hum.*

40: em_htgo_mus.*

41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
C 1	112	22.8	131704	2	AP004313	AP004313 Oryza sat
2	112	22.8	161809	2	AP005173	AP005173 Oryza sat
3	84.2	17.1	163381	2	AC128230	AC128230 Rattus no
4	78	15.9	177168	2	AC107145	AC107145 Rattus no
5	76.8	15.6	138467	2	AC111697	AC111697 Rattus no
6	74.4	15.1	82815	2	AC062001	AC062001 Homo sapi
7	73.2	14.9	158897	2	AC121439	AC121439 Rattus no
8	72.2	14.7	68330	2	AC116109	AC116109 Mus muscu
9	72.2	14.7	117082	2	AC126076	AC126076 Rattus no
C 10	70.4	14.3	101509	2	AC027353	AC027353 Homo sapi
11	69.8	14.2	151358	2	AC125906	AC125906 Rattus no
12	68.4	13.9	174605	2	AC121227	AC121227 Rattus no
13	68	13.8	174031	2	AC118303	AC118303 Rattus no
14	67.4	13.7	68787	2	AC106398	AC106398 Rattus no
15	67.4	13.7	130899	2	AC126730	AC126730 Rattus no
16	67.4	13.7	165988	2	AC121212	AC121212 Rattus no
17	67.2	13.7	185263	2	AC129765	AC129765 Rattus no
18	67.2	13.7	188231	2	AC094753	AC094753 Rattus no
19	66.8	13.6	179608	2	AC128497	AC128497 Rattus no
20	66.8	13.6	209216	2	AC117126	AC117126 Rattus no
C 21	66.6	13.5	99849	2	AC096824	AC096824 Rattus no
22	66.6	13.5	117082	2	AC126076	AC126076 Rattus no
C 23	66.6	13.5	118276	2	AP004120	AP004120 Oryza sat
24	66.6	13.5	158117	2	AP005111	AP005111 Oryza sat
C 25	66.6	13.5	158971	2	AP004882	AP004882 Oryza sat
C 26	66.6	13.5	165988	2	AC121212	AC121212 Rattus no
27	66.6	13.5	171624	2	AC095621	AC095621 Rattus no
28	66.2	13.5	194776	2	AC129706	AC129706 Rattus no
29	65.8	13.4	171204	2	AC130936	AC130936 Rattus no
30	65.6	13.3	175041	2	AC097133	AC097133 Rattus no
31	65.2	13.3	62839	2	AC098999	AC098999 Rattus no
32	65.2	13.3	170696	2	AP004658	AP004658 Oryza sat
C 33	65.2	13.3	171204	2	AC130936	AC130936 Rattus no
C 34	65	13.2	184245	2	AC127355	AC127355 Mus muscu
35	64.6	13.1	110000	2	AC116301_0	AC116301 Homo sapi
36	64.4	13.1	171852	2	AC058782	AC058782 Homo sapi
37	64.4	13.1	181022	2	AC121905	AC121905 Mus muscu
38	64.4	13.1	188133	2	AC112428	AC112428 Rattus no
39	64.2	13.0	81417	2	AC023265	AC023265 Homo sapi
40	64	13.0	69417	2	AC121516	AC121516 Mus muscu
41	64	13.0	162183	2	AC097753	AC097753 Rattus no
42	63.8	13.0	161163	2	AC115099	AC115099 Homo sapi
43	63.8	13.0	178941	2	AC113635	AC113635 Rattus no
C 44	63.6	12.9	129667	2	AC095300	AC095300 Rattus no
C 45	63.6	12.9	135119	2	AC011578	AC011578 Homo sapi

ALIGNMENTS

RESULT 1
AP004313/c
LOCUS AP004313 131704 bp DNA linear HTG 21-MAR-2002
DEFINITION Oryza sativa (japonica cultivar-group) chromosome 7 clone P0440B02,
*** SEQUENCING IN PROGRESS ***, in ordered pieces.
ACCESSION AP004313
VERSION HTG; HTGS_PHASE2.
KEYWORDS Oryza sativa (japonica cultivar-group) (cultivar:Nipponbare) DNA,
clone:P0440B02
SOURCE Oryza sativa (japonica cultivar-group)
ORGANISM Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.
REFERENCE 1

Sasaki, T., Matsumoto, T. and Yamamoto, K.
Oriza sativa nipponbare(GA3) genomic DNA, chromosome 7, PAC
clone:P0440802

JOURNAL
REFERENCE
PUBLISHED ONLY IN DATABASE (2001)

AUTHORS
TITLE
JOURNAL
DIRECT SUBMISSION
Sasaki, T., Matsumoto, T. and Yamamoto, K.

Submitted (07-NOV-2001) Takuji Sasaki, National Institute of
Agrobiological Sciences, Rice Genome Research Program; Kannondai
2-1-2, Tsukuba, Ibaraki 305-8602, Japan
(E-mail:tsasaki@nias.affrc.go.jp, URL:htc://rgp.dna.affrc.go.jp/,
Tel:81-298-38-7441, Fax:81-298-38-7468)

NOTE: It currently consists of 1 contigs. Gaps between the contigs
are represented as runs of N. The order of the pieces is believed
to be correct as given, however the sizes of the gaps between them
are based on estimates that have provided by the submitter. This
sequence will be replaced by the finished sequence as soon as it is
available and the accession number will be preserved.

- * NOTE: This is a 'working draft' sequence.
- * This sequence will be replaced
- * by the finished sequence as soon as it is available and
- * the accession number will be preserved.

FEATURES
source
Location/Qualifiers
1..131704
/organism="Oryza sativa (japonica cultivar-group)"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/chromosome="7"
/clone="P0440802"

BASE COUNT 38468 a 27101 c 28237 g 37798 t 100 others

ORIGIN

Query Match 22.8%; Score 112; DB 2: Length 131704;
Best Local Similarity 60.1%; Pred. No. 4.5e-11;
Matches 206; Conservative 0; Mismatches 131; Indels 6; Gaps 1;

```

QY      66  CCCCAATGCTCTTCCTTCTCCACCGACTCGGCCACTCGCCCTTCCCCGGCCTCG 125
        ||| ||||| ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 108777 CCAAAATGCTCCGGGCACCCTCCACCGCTCCACCCACTCGCCCTTCCCACCTCCCGCG 108718

QY      126  CCTCCGCAAAACCGCTTAACCCCGGTCTCGCGTCTCTCCGCTCCGGCGCGCTCGGCCCGTG 185
        ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| |||||
Db 108717 CTGCCCGCGCGCCCACTCCGGCAGGGCTTCTCACGGGTCTCGGCAGCGCGCGG 108658

QY      186  CGTCCGCGGTCTTAGCGCGAGTGGCGCCAGCGCATGAGTACGGTTAGGCCCTCGTTCG 245
        ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| |||||
Db 108657 TGCTTCGGGGGTGCGGGGTGACGACGGCGCACCGTCGGAGGATGGGCTTCGGCGGA 108598

Y       246  CCCTGCGGGCGCGCGCGGAGAGCTCGGCAGTGGG-----GGACTGCGCAGCGAGC 299
        ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| |||||
Db 108597 GCCCAGGCGAGCGGATGGAGGTGGCGATAGGAGGGGTAGGGTGGAGC 108538

QY      300  GGCTGTGGAGGTGGCGCAACGGCGCGGACGCTGCTGGGAGGTGCTCANGAAGTACT 359
        ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| |||||
Db 108537 GGCTAGTGGAGGTGGCGCAGGCGCGCAGCGCGCAGCGAGGTGCTCAGGAAGTACT 108478

QY      360  TCCGCCAGCGGGTTGAGATCATGCACAAGAGGACACAGTCC 402
        ||||| ||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 108477 TCCGGCAACCGCTCGAGATCATGCACAAGGAGGATCAGATGCC 108435

```

RESULT 2

AP005173

LOCUS AP005173 161809 bp DNA linear HTG 17-MAY-2002

DEFINITION Oryza sativa (japonica cultivar-group) chromosome 7 clone OSJNB0003E08, *** SEQUENCING IN PROGRESS ***, in ordered pieces.

ACCESSION AP005173

VERSION AP005173.1

KEYWORDS HTG: HTGS-PHASE2

SOURCE Oryza sativa (japonica cultivar-group) (cultivar:Nipponbare) DNA, clone:OSJNB0003E08.

ORGANISM Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

REFERENCE	AUTHORS	TITLE	JOURNAL	REFERENCE	AUTHORS	TITLE	JOURNAL
1	Sasaki, T., Matsumoto, T. and Katayose, Y.	Oryza sativa nipponbare(GA3) genomic DNA, chromosome 7, BAC clone:OSUNB00003E08	Published Only in Database (2002)				
2	(bases 1 to 161809)	Sasaki, T., Matsumoto, T. and Katayose, Y.	Direct Submision				
3	Submitted (15-MAY-2002)	Takuji Sasaki, National Institute of Agrobiological Sciences, Rice Genome Research Program; Kannondai 2-1-2, Tsukuba, Ibaraki 305-8602, Japan	(E-mail:tsasakienias.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/, Tel:81-298-38-7441, Fax:81-298-38-7468)				
4	NOTE: It currently consists of 1 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have provided by the submitter. This sequence will be replaced by the finished sequence as soon as it is available and the accession number will be preserved.	* NOTE: This is a 'working draft' sequence.	* This sequence will be replaced				
5	* by the finished sequence as soon as it is available and the accession number will be preserved.	* the accession number will be preserved.					
6	Location/Qualifiers						
7	1..161809	/organism="Oryza sativa (Japonica cultivar-group)"	/cultivar="Nipponbare"				
8	/db_xref="taxon:39947"	/chromosome="7"	/clone="OSJNB0003E08"				
9	BASE COUNT 46758 a 34058 c 34038 g 46734 t 211 others						
10	ORIGIN						
11	Query Match 22.8%; Score 112; DB 2: Length 161809;						
12	Best Local Similarity 60.1%; Pred. No. 4.4e-11;						
13	Matches 206; Conservative 0; Mismatches 131; Indels 6; Gaps 1;						
14	QY 66 CCCCAATGCTCTCTCTCTCCACCCACTCGGCCACTCGCCCTTCCCGGCTCG 135						
15							
16	Db 108667 CCCAATAGCTCCCGGCGACCTCCACCTGCCACCCACTCGCCCTTCCCGGCTCG 108726						
17							
18	QY 126 CCTCCGCAAAACCCCTAACCCCGCGTCTCGCTCTCTCCGCCCTCCGCGCCCTCGGCCGTGT 185						
19	Db 108727 CTGCCCCGCGCGCCACCTCCGCGCAGGGTCTCTCCACGGGTCCGTGGCGACGCGCGCG 108786						
20	QY 186 CGTCCGGGTCTTCAGCGCAGTGGCGGCCACCGCATGATACGCTTAGGGCCCTCGTTTCG 245						
21	Db 108787 TGGTTCGGGCGTCGGGGTTCACGAGGCCACACGCTCGCGAGATGGGCTCGGTCGTCGGGCGA 108846						
22	QY 246 CCCTGGGGGGGGCGCGCGAGAGCTCGCGCAGTGGG-----GGAGTTGGCGACGAGC 299						
23	Db 108847 GCCCAGGCGAGCGGATGGAGGTGCGGATAGGAGGGGTAGGGGTGAGC 108906						
24	QY 300 GGCTGTGGAGGTGGCCACGAGCGCGCGACGCTGTCTGGGAGGTGCTCANGAAGTACT 359						
25	Db 108907 GGCTAGTGGAGGTGGCGACAGGCGCGACACGCGCAGGGAGGTGCTCAGGAAGTACT 108966						
26	QY 360 TCCGCCAGCGGGTTGACATCATGCACAAAGGACACACATCC 402						
27	Db 108967 TCCGGCAACCGTCGATCATGCACAAAGGATGATGATGC 109009						
28	RESULT 3						
29	AC128230						
30	LOCUS	163381 bp DNA linear HTG 19-JUL-2002					
31	DEFINITION	Rattus norvegicus clone CH230-57K8, *** SEQUENCING IN PROGRESS ***;					
32		85 unordered pieces.					
33	AC128230						
34	ACCESSION	AC128230.1 GI:21908840					
35	VERSION	HTG; HTGS_PHASE1.					
36	KEYWORDS	Rattus norvegicus.					
37	SOURCE						

AC128230	AC128230	16381 bp	DNA	linear	HTG 19-JUL-2002
LOCUS	Rattus norvegicus clone CH230-57K8,	***	SEQUENCING IN PROGRESS	***,	
DEFINITION	85 unordered pieces.				
ACCESSION	AC128230				
VERSION	AC128230.1	GI:21908840			
KEYWORDS	HTG; HTGS_PHASE1.				
SOURCE	Rattus norvegicus.				

```

ORGANISM      Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 163381)
Muzny,D.M., Adams,C.C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Anaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
Barbieri,J., Benton,J., Binage,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Haylak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hognes,M., Holloway,C., Hollins,B.,
Homs,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moser,N., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokenkwo,S., Ogih,M., Okwuonu,G.,
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 163381)
Worley,K.C.
Direct Submission
Submitted (19-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GYBT
Center clone name: CH230-57K8
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 78315 bases at least Q40
Consensus quality: 87206 bases at least Q30
Consensus quality: 92805 bases at least Q20
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 85 contigs. The true order of the pieces

```

* is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

```

1 1027: contig of 1027 bp in length
1028 1127: gap of unknown length
1128 2195: contig of 1068 bp in length
2196 2295: gap of unknown length
2296 3623: contig of 1328 bp in length
3624 3723: gap of unknown length
3724 5144: contig of 1421 bp in length
5145 5245: gap of unknown length
5246 6262: contig of 1018 bp in length
6263 6363: gap of unknown length
6364 7887: contig of 1524 bp in length
7888 9581: contig of 1595 bp in length
9582 9681: gap of unknown length
9682 11192: contig of 1511 bp in length
11193 12292: gap of unknown length
12293 12398: contig of 1106 bp in length
12399 12498: gap of unknown length
12499 13696: contig of 1198 bp in length
13697 13796: gap of unknown length
13797 14976: contig of 1180 bp in length
14977 15076: gap of unknown length
15077 15562: contig of 1486 bp in length
15563 16663: gap of unknown length
16664 17680: contig of 1018 bp in length
17681 17781: gap of unknown length
17782 18800: contig of 1020 bp in length
18801 18900: gap of unknown length
18901 20003: contig of 1103 bp in length
20004 21681: contig of 1578 bp in length
21682 21781: gap of unknown length
21782 22870: contig of 1089 bp in length
22871 22970: gap of unknown length
22971 24090: contig of 1120 bp in length
24091 24190: gap of unknown length
24191 25426: contig of 1236 bp in length
25427 25527: gap of unknown length
25528 26690: contig of 1163 bp in length
26691 28356: contig of 1567 bp in length
28357 28456: gap of unknown length
28457 30062: contig of 1606 bp in length
30063 30162: gap of unknown length
30163 31306: contig of 1144 bp in length
31307 31406: gap of unknown length
31407 32968: contig of 1562 bp in length
32969 33068: gap of unknown length
33069 34291: contig of 1223 bp in length
34292 34391: gap of unknown length
34392 34667: contig of 1076 bp in length
34668 35567: gap of unknown length
35568 36788: contig of 1221 bp in length
36789 36888: gap of unknown length
36889 38584: contig of 1696 bp in length
38585 38684: gap of unknown length
38685 40070: contig of 1386 bp in length
40071 40170: gap of unknown length
40171 41693: contig of 1523 bp in length
41694 41793: gap of unknown length
41794 42956: contig of 1163 bp in length
42957 43056: gap of unknown length
43057 44652: contig of 1596 bp in length
44653 44753: gap of unknown length
44754 46313: contig of 1567 bp in length
46314 46419: gap of unknown length
46420 48471: contig of 2052 bp in length

```



```

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: G00J
Center clone name: CH230-106E18
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 120304 bases at least Q40
Consensus quality: 126448 bases at least Q30
Consensus quality: 131535 bases at least Q20
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 67 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1 1020: contig of 1020 bp in length
* 1021: contig of unknown length
* 1121: 2478: contig of 1358 bp in length
* 2479: 2578: gap of unknown length
* 2579: 3589: contig of 1011 bp in length
* 3590: 3689: gap of unknown length
* 3690: 4826: contig of 1137 bp in length
* 4827: 4926: gap of unknown length
* 4927: 6143: contig of 1217 bp in length
* 6144: 6243: gap of unknown length
* 6244: 7578: contig of 1335 bp in length
* 7579: 7678: gap of unknown length
* 7679: 8692: contig of 1014 bp in length
* 8693: 8792: gap of unknown length
* 8793: 9896: contig of 1104 bp in length
* 9897: 9996: gap of unknown length
* 9997: 11336: contig of 1340 bp in length
* 11337: 11436: gap of unknown length
* 11437: 13217: contig of 1781 bp in length
* 13218: 13317: gap of unknown length
* 13318: 15162: contig of 1845 bp in length
* 15163: 15262: gap of unknown length
* 15263: 16473: contig of 1211 bp in length
* 16474: 16573: gap of unknown length
* 16574: 17382: contig of 1009 bp in length
* 17583: 17682: gap of unknown length
* 17683: 18900: contig of 1218 bp in length
* 18901: 19000: gap of unknown length
* 19001: 20997: contig of 1097 bp in length
* 20997: 20197: gap of unknown length
* 20998: 20197: gap of unknown length
* 20198: 21776: contig of 1579 bp in length
* 21777: 21876: gap of unknown length
* 21877: 23292: contig of 1416 bp in length
* 23292: 23392: gap of unknown length
* 23393: 24939: contig of 1547 bp in length
* 24940: 25039: gap of unknown length
* 25040: 27311: contig of 2272 bp in length
* 27312: 27411: gap of unknown length
* 27412: 28968: contig of 1557 bp in length
* 28969: 29068: gap of unknown length
* 29069: 30742: contig of 1674 bp in length
* 30743: 30842: gap of unknown length
* 30843: 32409: contig of 1567 bp in length
* 32410: 32509: gap of unknown length
* 32510: 34377: contig of 1868 bp in length
* 34378: 34477: gap of unknown length
* 34478: 35582: contig of 1105 bp in length

```

```
Qy 139 TAACCCCGGCTCTCGCTCTCCCTCCGCGCGCGCTCCGCGCTCGCTCCGCGGTCTT 198
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 81438 CNNNNCCNNNNCCNNNNCCNNNNCCNNNNCCNNNNCCNNNNCCNNNNCCNNNNCCNN 81497
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 199 GAGCGCGAGTGGCGGACGAGTACGATAGGTAGGCTCTGTCGCCCTGGGCGGCG 258
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 81498 GCGCGCGGNGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 81557
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 259 CGCGCGGAGAGCTCGCGAGTGGGCGAGTGGGCGAGCGAGCGCTGGTGGAGGTGGCGCA 318
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 81558 CGCGCGGCGGNGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 81617
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 319 ACGGCGGCGGACGCTCTGCTGGGAGGTGCTC 349
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 81618 CCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 81648
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 5
AC111697
*****
AC111697
AC111697 138467 bp DNA linear HTG 13-JUL-2002
Rattus norvegicus clone CH230-20C20, *** SEQUENCING IN PROGRESS
***, 59 unordered pieces.
AC111697
HTG: HTGS_PHASE1.
Norway rat.
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 138467)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-olsman,F.R., Allen,C.,
Alsbrooks,S.L., Anaratinge,H.C., Are,J.R., Ayele,M., Banks,T.,
Barbieri,J., Benton,J., Binage,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carroll,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hogue,M., Holloway,C., Hollins,B.,
Homs,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karlisson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louissegh,H.,
Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokkwo,S., Ogih,M., Okwuonu,G.,
Oragune,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Plckens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
Rives,M., Rojas,A., Rojokan,I., Rolfe,M., Ruiz,S., Savery,G.,
Scherer,S., Scott,G., Shen,H., Shoohtari,N., Sisson,I.,
Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatek,A., Taber,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Usmani,K., Vasquez,L., Vera,V., Villalobos,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.
Direct Submission
Unpublished
```

REFERENCE
AUTHORS
TITLE
JOURNALREFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT

2 (bases 1 to 138467)
Worley,K.C.
Direct Submission
Submitted (19-FEB-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 138467)
Worley,K.C.
Direct Submission
Submitted (13-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 12, 2002 this sequence version replaced gi:18701464.

Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GMWR
Center clone name: CH230-20C20
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator; Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 97272 bases at least Q40
Consensus quality: 104545 bases at least Q30
Consensus quality: 110146 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
consists of 59 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.

1 1060: contig of 1060 bp in length
1061 60: gap of unknown length
1161 28: contig of 1068 bp in length
2229 2328: gap of unknown length
2329 3520: contig of 1192 bp in length
3521 3620: gap of unknown length
3621 4797: contig of 1177 bp in length
4798 4897: gap of unknown length
4898 6624: contig of 1727 bp in length
6625 6724: gap of unknown length
6725 8301: contig of 1577 bp in length
8302 8401: gap of unknown length
8402 9813: contig of 1412 bp in length
9814 9913: gap of unknown length
9914 11128: contig of 1215 bp in length
11129 12658: gap of unknown length
12659 12758: contig of 1430 bp in length
12759 15360: contig of 2602 bp in length
15361 15461: gap of unknown length
15461 17026: contig of 1565 bp in length
17026 17126: gap of unknown length
17126 18259: contig of 1134 bp in length
18260 18359: gap of unknown length
18360 19979: contig of 1620 bp in length
19980 20079: gap of unknown length
20080 21208: contig of 1129 bp in length
21209 21308: gap of unknown length
21309 22587: contig of 1279 bp in length
22588 22688: gap of unknown length
22688 24387: contig of 1700 bp in length
24388 24488: gap of unknown length
24488 26131: contig of 1644 bp in length
26131 26231: gap of unknown length

* 26232 27699: contig of 1468 bp in length
* 27700 gap of unknown length
* 27800 29145: contig of 1346 bp in length
* 29246 29245: gap of unknown length
* 30338 30437: gap of unknown length
* 30438 31510: contig of 1073 bp in length
* 31511 31610: gap of unknown length
* 31611 33147: contig of 1537 bp in length
* 33148 33247: gap of unknown length
* 33248 34743: contig of 1496 bp in length
* 34744 34843: gap of unknown length
* 34844 36324: contig of 1481 bp in length
* 36325 36424: gap of unknown length
* 36425 37988: contig of 1564 bp in length
* 37989 38088: gap of unknown length
* 38089 40522: contig of 2434 bp in length
* 40523 40622: gap of unknown length
* 40624 41877: contig of 1254 bp in length
* 41878 41976: gap of unknown length
* 41977 43465: contig of 1489 bp in length
* 43466 43565: gap of unknown length
* 43566 45265: contig of 1700 bp in length
* 45266 45365: gap of unknown length
* 45366 47618: contig of 2253 bp in length
* 47619 47718: gap of unknown length
* 47719 49624: contig of 1906 bp in length
* 49625 49724: gap of unknown length
* 49725 51701: contig of 1977 bp in length
* 51702 51801: gap of unknown length
* 51802 54174: contig of 2373 bp in length
* 54175 54274: gap of unknown length
* 54275 56347: contig of 2073 bp in length
* 56348 56447: gap of unknown length
* 56448 57737: contig of 1290 bp in length
* 57738 57837: gap of unknown length
* 57838 60324: contig of 2487 bp in length
* 60325 60425: gap of unknown length
* 60426 61731: contig of 1307 bp in length
* 61732 61831: gap of unknown length
* 61832 63545: contig of 1714 bp in length
* 63546 63645: gap of unknown length
* 63646 65262: contig of 1617 bp in length
* 65263 65362: gap of unknown length
* 65363 67811: contig of 2449 bp in length
* 67812 67911: gap of unknown length
* 67912 70170: contig of 2259 bp in length
* 70171 70270: gap of unknown length
* 70271 72210: contig of 1940 bp in length
* 72211 72310: gap of unknown length
* 72311 73697: contig of 1387 bp in length
* 73698 73797: gap of unknown length
* 73798 75398: contig of 1601 bp in length
* 75399 75498: gap of unknown length
* 75499 78702: contig of 3204 bp in length
* 78703 78802: gap of unknown length
* 78803 81340: contig of 2538 bp in length
* 81341 81440: gap of unknown length
* 81441 84907: contig of 3467 bp in length
* 84908 85007: gap of unknown length
* 85008 88196: contig of 3189 bp in length
* 88197 88296: gap of unknown length
* 88297 91902: contig of 3606 bp in length
* 91903 92002: gap of unknown length
* 92003 95117: contig of 3115 bp in length
* 95118 95217: gap of unknown length
* 95218 98243: contig of 3026 bp in length
* 98244 98343: gap of unknown length
* 98344 102722: contig of 4379 bp in length
* 102723 102822: gap of unknown length
* 102823 105855: contig of 3033 bp in length
* 105856 105956: gap of unknown length
* 110171: contig of 4216 bp in length

Query Match 15.6%; Score 76.8; DB 2; Length 138467;
Best Local Similarity 52.4%; Pred. No. 0.00013;
Matches 176; Conservative 0; Mismatches 159; Indels 1; Gaps 1;
QY 17 CTCGGCCCGGACATTAAACCTTGTCTCCCGACAAACCGCGCCGACTCTCTCCCAATGCTC 76
DB 2848 CNGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCC 2907
QY 77 TCCTCTTCCTCTCCACCCACTCGCGCACCTCGCCCTTCCCGCGGCTCGCTCCGCAAAAC 136
DB 2908 CTCTCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCC 2967
QY 137 CCTAACCCCGGCTCTCGCTCTCTCGGCTCGCGCGCGCGCTCGCGCTGCTCGCGGTC 196
DB 2968 CCGCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCGCGCGCGC 3027
QY 197 TTGAGCGCGAGTGGCGCGCCAGCCGATGATGATGATGATGATGATGATGATGATGATG 256
DB 3028 CGC 3087
QY 257 GCGCGCGGAGAGCTCGCGGAGTGGGGAGTGGCGACGAGGAGGCTGCTGAGGAGTGGCG 316
DB 3088 GC 3146
QY 317 CAACGGC 352
DB 3147 GC 3182

RESULT 6
AC062001
LOCUS
DEFINITION
AC062001
VERSION
HTG: HTGS_PHASE0.
SOURCE
Homo sapiens.
ORGANISM
Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 82815)
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N., Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F., Boguslavsky, L., Boukhgalter, B., Brown, A., Burkett, G., Campopiano, A., Castle, A., Choe, P., Cho, P., Colangelo, M., Collins, S., Collymore, A., Cooke, P., DeArrellano, K., Dewar, K., Diaz, J.S., Dodge, S., Domino, M., Doyle, M., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gargano, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L., Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., LaRocque, K., Lamazares, R., Landers, T., Lechoczy, J., Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N., McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheters, R., Meldrum, J., Meneus, L., Mihova, T., Miranda, C., Mlenka, V., Morrow, J., Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, T.M., Oliver, J., Peterson, K., Pierre, N., Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Triggillo, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zimmer, A. and Zody, M.
Direct Submission
Submitted (21-APR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L9871
Center clone name: 124_R_2

* NOTE: This record contains 78 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1 975: contig of 975 bp in length
* 976 1075: gap of 100 bp
* 1076 2100: contig of 1025 bp in length
* 2101 2200: gap of 100 bp
* 2201 3185: contig of 985 bp in length
* 3186 3285: gap of 100 bp
* 3286 4151: contig of 866 bp in length
* 4152 4251: gap of 100 bp
* 4252 5206: contig of 955 bp in length
* 5207 5306: gap of 100 bp
* 5307 6287: contig of 981 bp in length
* 6288 6387: gap of 100 bp
* 6388 7332: contig of 945 bp in length
* 7333 7432: gap of 100 bp
* 7433 8397: contig of 965 bp in length
* 8398 8497: gap of 100 bp
* 8498 9457: contig of 960 bp in length
* 9458 9557: gap of 100 bp
* 9558 10489: contig of 932 bp in length
* 10490 10589: gap of 100 bp
* 10590 11602: contig of 1013 bp in length
* 11603 11702: gap of 100 bp
* 11703 12694: contig of 992 bp in length
* 12695 12794: gap of 100 bp
* 12795 13798: contig of 1004 bp in length
* 13799 13898: gap of 100 bp
* 13899 14866: contig of 968 bp in length
* 14867 14966: gap of 100 bp
* 14967 15930: contig of 964 bp in length
* 15931 16030: gap of 100 bp
* 16031 16981: contig of 951 bp in length
* 16982 17081: gap of 100 bp
* 17082 18043: contig of 962 bp in length
* 18044 18143: gap of 100 bp
* 18144 19119: contig of 976 bp in length
* 19120 19219: gap of 100 bp
* 19220 20173: contig of 954 bp in length
* 20174 20273: gap of 100 bp
* 20274 21232: contig of 959 bp in length
* 21233 21332: gap of 100 bp
* 21333 22227: contig of 895 bp in length
* 22228 23227: gap of 100 bp
* 23228 23303: contig of 976 bp in length
* 23304 23403: gap of 100 bp
* 23404 24393: contig of 990 bp in length
* 24394 24493: gap of 100 bp
* 24494 25495: contig of 1002 bp in length
* 25496 25595: gap of 100 bp
* 25596 26554: contig of 959 bp in length
* 26555 26654: gap of 100 bp
* 26655 27611: contig of 957 bp in length
* 27612 27711: gap of 100 bp

27712 28689: contig of 978 bp in length
* 28690 28789: gap of 100 bp
* 28790 29750: contig of 961 bp in length
* 29751 29850: gap of 100 bp
* 29851 30790: contig of 940 bp in length
* 30791 30890: gap of 100 bp
* 30891 31881: contig of 991 bp in length
* 31882 31981: gap of 100 bp
* 31982 32962: contig of 981 bp in length
* 32963 33062: gap of 100 bp
* 33063 34046: contig of 984 bp in length
* 34047 34146: gap of 100 bp
* 34147 35108: contig of 962 bp in length
* 35109 35208: gap of 100 bp
* 35209 36136: contig of 928 bp in length
* 36137 36236: gap of 100 bp
* 36237 37202: contig of 966 bp in length
* 37203 37302: gap of 100 bp
* 37303 38272: contig of 970 bp in length
* 38273 38372: gap of 100 bp
* 38373 39319: contig of 947 bp in length
* 39320 39419: gap of 100 bp
* 39420 40376: contig of 957 bp in length
* 40377 40476: gap of 100 bp
* 40477 41464: contig of 988 bp in length
* 41465 41564: gap of 100 bp
* 41565 42492: contig of 928 bp in length
* 42493 42592: gap of 100 bp
* 42593 43578: contig of 986 bp in length
* 43579 43678: gap of 100 bp
* 43679 44686: contig of 1008 bp in length
* 44687 44786: gap of 100 bp
* 44787 45745: contig of 959 bp in length
* 45746 45845: gap of 100 bp
* 45846 46827: contig of 982 bp in length
* 46828 46927: gap of 100 bp
* 46928 47910: contig of 983 bp in length
* 47911 48010: gap of 100 bp
* 48011 48969: contig of 959 bp in length
* 48970 49069: gap of 100 bp
* 49070 50028: contig of 959 bp in length
* 50029 50128: gap of 100 bp
* 50129 51053: contig of 925 bp in length
* 51054 51153: gap of 100 bp
* 51154 52111: contig of 938 bp in length
* 52112 52211: gap of 100 bp
* 52212 53189: contig of 978 bp in length
* 53190 53289: gap of 100 bp
* 53290 54271: contig of 982 bp in length
* 54272 54371: gap of 100 bp
* 54372 55146: contig of 965 bp in length
* 55337 55436: gap of 100 bp
* 55437 56366: contig of 930 bp in length
* 56367 56466: gap of 100 bp
* 56467 57415: contig of 949 bp in length
* 57416 57515: gap of 100 bp
* 57516 58501: contig of 986 bp in length
* 58502 58601: gap of 100 bp
* 58602 59567: contig of 966 bp in length
* 59568 59667: gap of 100 bp
* 59668 60585: contig of 918 bp in length
* 60586 60685: gap of 100 bp
* 60686 61646: contig of 961 bp in length
* 61647 61746: gap of 100 bp
* 61747 62739: contig of 993 bp in length
* 62740 62839: gap of 100 bp
* 62840 63824: contig of 985 bp in length
* 63825 63924: gap of 100 bp
* 63925 64903: contig of 979 bp in length
* 64904 65003: gap of 100 bp
* 65004 65963: contig of 960 bp in length
* 65964 66063: gap of 100 bp
* 66064 67010: contig of 947 bp in length

1 (bases 1 to 158897)

REFERENCE
AUTHORS

Muzny, D.M., Adams, C.C., Adio-Oduola, B., Ali-osman, F.R., Allen, C.,
Alsbrooks, S.L., Anaratunge, H.C., Are, J.R., Ayele, M., Banks, T.,
Barack, J., Benton, J., Binage, K., Blankenburg, K., Bonnin, D.,
Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P.,
Buahy, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,
Carton, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,
Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,
Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,
Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,
DeLaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,
Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,
Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,
Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,
Gabisai, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,
Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,
Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J.,

COMMENT

[illegible]

TITLE
JOURNAL
REFERENCE
AUTHORS

Mus musculus, clone RP23-467G14
Unpublished
2 (bases 1 to 68330)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L.,
Boukigalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
Landers,T., Lehoczy,J., Levine,R., Lindblad-Toh,K., Liu,G.,
MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., Meldrum,J., Meneus,L.,
Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Olivier,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trifilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (25-MAR-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RW/RepeatMasker.html>
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence.submissions@genome.wi.mit.edu
----- Project Information
Center project name: L22810
Center clone name: 467_G14

TITLE
JOURNAL
COMMENT

* NOTE: This record contains 85 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1 647: contig of 647 bp in length
* 648 747: gap of 100 bp
* 748 1422: contig of 675 bp in length
* 1423 1522: gap of 100 bp
* 1523 2229: contig of 707 bp in length
* 2230 2329: gap of 100 bp
* 2330 3001: contig of 672 bp in length
* 3002 3101: gap of 100 bp
* 3102 3783: contig of 682 bp in length
* 3784 3883: gap of 100 bp
* 3884 4586: contig of 703 bp in length
* 4587 4686: gap of 100 bp
* 4687 5381: contig of 695 bp in length
* 5382 5481: gap of 100 bp
* 5482 6190: contig of 709 bp in length
* 6191 6290: gap of 100 bp
* 6291 7003: contig of 713 bp in length
* 7004 7103: gap of 100 bp
* 7104 7824: contig of 721 bp in length
* 7825 7924: gap of 100 bp
* 7925 8662: contig of 738 bp in length
* 8663 8762: gap of 100 bp
* 8763 9402: contig of 640 bp in length
* 9403 9502: gap of 100 bp
* 9503 10235: contig of 733 bp in length
* 10236 10335: gap of 100 bp
* 10336 11055: contig of 720 bp in length
* 11056 11155: gap of 100 bp
* 11156 11861: contig of 706 bp in length
* 11862 11961: gap of 100 bp
* 11962 12853: contig of 692 bp in length
* 12854 12753: gap of 100 bp
* 12754 13445: contig of 692 bp in length
* 13446 13545: gap of 100 bp
* 13546 14235: contig of 690 bp in length
* 14236 14335: gap of 100 bp
* 14336 15046: contig of 711 bp in length
* 15047 15146: gap of 100 bp
* 15147 15853: contig of 707 bp in length
* 15854 15953: gap of 100 bp
* 15954 16632: contig of 679 bp in length
* 16633 16732: gap of 100 bp
* 16733 17441: contig of 709 bp in length
* 17442 17541: gap of 100 bp
* 17542 18270: contig of 729 bp in length
* 18271 18370: gap of 100 bp
* 18371 19048: contig of 678 bp in length
* 19049 19148: gap of 100 bp
* 19149 19851: contig of 703 bp in length
* 19852 19951: gap of 100 bp
* 19952 20661: contig of 710 bp in length
* 20662 20761: gap of 100 bp
* 20762 21473: contig of 712 bp in length
* 21474 21573: gap of 100 bp
* 21574 22286: contig of 713 bp in length
* 22287 22386: gap of 100 bp
* 22387 23086: contig of 700 bp in length
* 23087 23186: gap of 100 bp
* 23187 23920: contig of 734 bp in length
* 23921 24020: gap of 100 bp
* 24021 24713: contig of 693 bp in length
* 24714 24813: gap of 100 bp
* 24814 25440: contig of 727 bp in length
* 25441 25641: gap of 100 bp
* 25641 26364: contig of 724 bp in length
* 26365 26464: gap of 100 bp
* 26465 27180: contig of 716 bp in length
* 27181 27280: gap of 100 bp
* 27281 27977: contig of 697 bp in length
* 27978 28077: gap of 100 bp
* 28078 28769: contig of 692 bp in length
* 28770 28869: gap of 100 bp
* 28870 29559: contig of 690 bp in length
* 29560 29659: gap of 100 bp
* 29660 30357: contig of 698 bp in length
* 30358 30457: gap of 100 bp
* 30458 31165: contig of 708 bp in length
* 31166 31265: gap of 100 bp
* 31266 31948: contig of 683 bp in length
* 31949 32048: gap of 100 bp
* 32049 32768: contig of 720 bp in length
* 32769 32868: gap of 100 bp
* 32869 33548: contig of 680 bp in length
* 33549 33648: gap of 100 bp
* 33649 34357: contig of 709 bp in length
* 34358 34457: gap of 100 bp
* 34458 35187: contig of 730 bp in length
* 35188 35287: gap of 100 bp
* 35288 35990: contig of 703 bp in length
* 35991 36090: gap of 100 bp
* 36091 36767: contig of 677 bp in length
* 36768 36867: gap of 100 bp
* 36868 37575: contig of 708 bp in length
* 37576 37675: gap of 100 bp
* 37676 38379: contig of 704 bp in length


```
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 55139 bases at least Q40
Consensus quality: 61060 bases at least Q30
Consensus quality: 64630 bases at least Q20
-----
* NOTE: Estimated insert size may differ from sequence length
      (see http://www.hgsc.bcm.tmc.edu/docs/genbank\_draft\_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 63 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1
* 1055: contig of 1055 bp in length
* 1056: 1155: gap of unknown length
* 1156: 2165: contig of 1010 bp in length
* 2165: 2265: gap of unknown length
* 2265: 3314: contig of 1049 bp in length
* 3314: 3414: gap of unknown length
* 3414: 5018: contig of 1604 bp in length
* 5018: 5118: gap of unknown length
* 5118: 6243: contig of 1125 bp in length
* 6243: 6343: gap of unknown length
* 6343: 7662: contig of 1319 bp in length
* 7662: 7763: gap of unknown length
* 7763: 9164: contig of 1402 bp in length
* 9164: 9264: gap of unknown length
* 9264: 10353: contig of 1088 bp in length
* 10353: 10452: gap of unknown length
* 10452: 11637: contig of 1185 bp in length
* 11637: 11737: gap of unknown length
* 11737: 13085: contig of 1348 bp in length
* 13085: 13185: gap of unknown length
* 13185: 14617: contig of 1432 bp in length
* 14617: 14717: gap of unknown length
* 14717: 16407: contig of 1690 bp in length
* 16407: 16507: gap of unknown length
* 16507: 17676: contig of 1169 bp in length
* 17676: 17776: gap of unknown length
* 17776: 19207: contig of 1431 bp in length
* 19207: 19307: gap of unknown length
* 19307: 21017: contig of 1710 bp in length
* 21017: 21117: gap of unknown length
* 21117: 22172: contig of 1055 bp in length
* 22172: 22722: gap of unknown length
* 22722: 23366: contig of 1094 bp in length
* 23366: 23466: gap of unknown length
* 23466: 24549: contig of 1083 bp in length
* 24549: 24649: gap of unknown length
* 24649: 25665: contig of 1016 bp in length
* 25665: 25765: gap of unknown length
* 25765: 27421: contig of 1636 bp in length
* 27421: 27521: gap of unknown length
* 27521: 28564: contig of 1043 bp in length
* 28564: 28664: gap of unknown length
* 28664: 29755: contig of 1091 bp in length
* 29755: 29855: gap of unknown length
* 29855: 31857: contig of 1802 bp in length
* 31857: 31757: gap of unknown length
* 31757: 33595: contig of 1838 bp in length
* 33595: 33695: gap of unknown length
* 33695: 35198: contig of 1503 bp in length
* 35198: 35298: gap of unknown length
* 35298: 36466: contig of 1168 bp in length
* 36466: 36566: gap of unknown length
* 36566: 37640: contig of 1074 bp in length
* 37640: 37741: gap of unknown length
* 37741: 39003: contig of 1263 bp in length
*

```

```

39004
39104
41305: gap of unknown length
41305: contig of 2202 bp in length
41306: 41405: gap of unknown length
41405: contig of 1215 bp in length
42620: gap of unknown length
42621: 42720: gap of unknown length
42721: 43940: contig of 1220 bp in length
43941: 44040: gap of unknown length
44041: 45996: contig of 1956 bp in length
45997: 46096: gap of unknown length
46097: 47529: contig of 1433 bp in length
47530: 47629: gap of unknown length
47630: 49451: contig of 1822 bp in length
49452: 49551: gap of unknown length
49552: 50801: contig of 1250 bp in length
50802: 50901: gap of unknown length
50902: 52881: contig of 1980 bp in length
52882: 52981: gap of unknown length
52982: 54499: contig of 1518 bp in length
54499: 54599: gap of unknown length
54600: 55934: contig of 1335 bp in length
55935: 56034: gap of unknown length
56035: 58138: contig of 2104 bp in length
58139: 58238: gap of unknown length
58239: 59732: contig of 1494 bp in length
59733: 59832: gap of unknown length
59833: 61285: contig of 1453 bp in length
61286: 61385: gap of unknown length
61386: 62880: contig of 1495 bp in length
62881: 64930: contig of 1950 bp in length
64931: 65030: gap of unknown length
65031: 66998: contig of 1968 bp in length
66999: 67098: gap of unknown length
67099: 69768: contig of 2670 bp in length
69769: 69868: gap of unknown length
69869: 72122: contig of 2254 bp in length
72123: 72222: gap of unknown length
72223: 74144: contig of 1922 bp in length
74145: 74244: gap of unknown length
74245: 75957: contig of 1713 bp in length
75958: 76057: gap of unknown length
76058: 78577: contig of 2520 bp in length
78578: 78677: gap of unknown length
78678: 80255: contig of 1578 bp in length
80256: 80355: gap of unknown length
80356: 82593: contig of 2238 bp in length
82594: 84693: gap of unknown length
84694: 84785: contig of 1992 bp in length
84786: 86448: contig of 1663 bp in length
86449: 87884: gap of unknown length
87885: 87884: contig of 1336 bp in length

Query Match
Best Local Similarity 14.7%; Score 72.2; DB 2: Length 117082;
Matches 152; Conservative 0; Mismatches 189; Indels 0; Gaps 0;

QY 9 GCGTCATTCTCGGCCCGCCGACATTTAAACCTTCCTCCGACACACCGCCGCGGCTCTCC 58
DB 48749 GNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 48808
QY 69 CAATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 128
DB 48809 CNNAAACCCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCC 48868
QY 129 CCGCAAAACCTTAACCCCGCGCTCTCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 188
DB 48869 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 48928
QY 189 CCGCGGCTTGTGAGCGGAGTGCGGCCAGCGGATGAGTACGCTTACGCGCTCTCTCTCG 248
DB 48929 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 48988

```



```
* 56357 57447: contig of 1091 bp in length
* 57448 57547: gap of unknown length
* 57548 58710: contig of 1163 bp in length
* 58711 59999: gap of unknown length
* 58810 60099: contig of 1189 bp in length
* 60000 61343: gap of unknown length
* 61344 61444: contig of 1244 bp in length
* 61444 62568: gap of unknown length
* 62569 62688: contig of 1125 bp in length
* 62689 63904: gap of unknown length
* 63905 64004: contig of 1236 bp in length
* 64005 65045: gap of unknown length
* 65046 66360: contig of 1041 bp in length
* 66361 66460: gap of unknown length
* 66461 67705: contig of 1215 bp in length
* 67706 68914: gap of unknown length
* 68915 69014: contig of 1245 bp in length
* 69015 70199: contig of 1109 bp in length
* 70200 70299: gap of unknown length
* 70300 71355: contig of 1185 bp in length
* 71356 71455: gap of unknown length
* 71456 73215: contig of 1056 bp in length
* 73216 73315: gap of unknown length
* 73316 74456: contig of 1760 bp in length
* 74457 74556: gap of unknown length
* 74557 76312: contig of 1141 bp in length
* 76313 76412: gap of unknown length
* 76413 77484: contig of 1756 bp in length
* 77485 77584: gap of unknown length
* 77585 78684: contig of 1072 bp in length
* 78685 78783: gap of unknown length
* 78784 79558: contig of 1099 bp in length
* 79559 79958: gap of unknown length
* 79959 80987: contig of 1075 bp in length
* 80988 81087: gap of unknown length
* 81088 82188: contig of 1029 bp in length
* 82189 82289: contig of 1101 bp in length
* 82290 83320: gap of unknown length
* 83321 83420: contig of 1032 bp in length
* 83421 84521: gap of unknown length
* 84522 84621: contig of 1101 bp in length
* 84622 101509: gap of unknown length
* 101509: contig of 16888 bp in length.

FEATURES
  source
    1. .101509
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /chromosome="16"
      /clone="RP11-167D21"
      /clone_lib="RPC1 human BAC library 11"
BASE COUNT 27989 a 7533 c 28794 g 7012 t 30181 others
ORIGIN

Query Match 14.3%; Score 70.4; DB 2; Length 101509;
Best Local Similarity 43.3%; Pred. No. 0.002;
Matches 143; Conservative 0; Mismatches 187; Indels 0; Gaps 0;

QY 17 CTCGGCCCGGACATTTAAACCTTGTCTCGCCGACAAACCGCGCGGACCTCTCCCAATGCTC 76
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 31049 CNCNCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCC 30990
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 77 TCCTCTCTCTCTCAGCCACTGGCCGACCTGCGCTTTCCTCCCGGCTCGCGCTCCGCAAAAC 136
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 30989 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 30930
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 137 CCTAACCGCCGCTCTCGCTCTCTCGCCCTCGCGCGCGCGCGCGCTGCTGCTCGCGGTC 196
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 30929 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 30870
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 197 TTGAGCGCGAGTGGGCGCGCGCGCGATGAGTACGGTTAGGGCCCTCGTTCGCGCGTGGGGG 256
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```


Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,
 Nguyen, N., Nickerson, E., Nwokkwo, S., Oguh, M., Okwuonu, G.,
 Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,
 Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y.,
 Rives, M., Rojas, A., Rojubokan, I., Rolfe, M., Ruiz, S., Savary, G.,
 Scherer, S., Scott, G., Shen, H., Shooshtari, N., Sisson, I.,
 Sodergren, E., Sonaike, T., Sparks, A., Stanley, H., Stone, H.,
 Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,
 Tansley, J., Taylor, C., Taylor, T., Teifrod, B., Thomas, N., Thomas, S.,
 Umanu, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wang, S.,
 Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,
 Williams, G., Williamson, A., Wleczyk, R., Woodden, S., Worley, K.,
 Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
 Weinstock, G. and Gibbs, R.
 Direct Submission
 Unpublished
 2 (bases 1 to 174031)
 Direct Submission
 Worley, K.C.
 Direct Submission
 Submitted (15-APR-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 174031)
 Direct Submission
 Worley, K.C.
 Direct Submission
 Submitted (18-JUL-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Jul 14, 2002 this sequence version replaced gl:20330629.

 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu

 Project Information
 Center project name: GUJG
 Center clone name: CH230-173H17

 Summary Statistics
 Sequencing vector: Plasmid
 Chemistry: Dye-terminator Big Dye: 100% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 107398 bases at least Q40
 Consensus quality: 119658 bases at least Q30
 Consensus quality: 127269 bases at least Q20

 * NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 58 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 * 1 1024: contig of 1024 bp in length
 * 1025 1124: gap of unknown length
 * 1125 2167: contig of 1043 bp in length
 * 2168 2267: gap of unknown length
 * 2268 3414: contig of 1147 bp in length
 * 3415 3514: gap of unknown length
 * 3515 4714: contig of 1200 bp in length
 * 4715 4814: gap of unknown length
 * 4815 6353: contig of 1539 bp in length
 * 6354 6453: gap of unknown length
 * 6454 7915: contig of 1462 bp in length
 * 7916 8015: gap of unknown length
 * 8016 9359: contig of 1344 bp in length
 * 9360 9459: gap of unknown length
 * 9460 11082: contig of 1623 bp in length
 * 11083 11182: gap of unknown length
 * 11183 12369: contig of 1187 bp in length
 * 12370 12469: gap of unknown length
 * 12470 13930: contig of 1461 bp in length
 * 13931 14030: gap of unknown length
 * 14031 15253: contig of 1223 bp in length
 * 15254 15353: gap of unknown length
 * 15354 16696: contig of 1343 bp in length
 * 16697 16796: gap of unknown length
 * 16797 18403: contig of 1607 bp in length
 * 18404 18503: gap of unknown length
 * 18504 19776: contig of 1273 bp in length
 * 19777 19876: gap of unknown length
 * 19877 21648: contig of 1772 bp in length
 * 21649 23044: gap of unknown length
 * 23045 23194: gap of unknown length
 * 23195 25192: contig of 1898 bp in length
 * 25193 26802: contig of 1610 bp in length
 * 26803 26902: gap of unknown length
 * 26903 28826: contig of 1824 bp in length
 * 28827 31064: contig of 2238 bp in length
 * 31065 31164: gap of unknown length
 * 31165 33254: contig of 2090 bp in length
 * 33255 33535: gap of unknown length
 * 33536 35764: gap of unknown length
 * 35765 36989: contig of 1126 bp in length
 * 36990 37089: gap of unknown length
 * 37090 38703: contig of 1614 bp in length
 * 38704 38803: gap of unknown length
 * 38804 40934: contig of 2131 bp in length
 * 40935 41034: gap of unknown length
 * 41035 42978: contig of 1944 bp in length
 * 42979 43078: gap of unknown length
 * 43079 45214: contig of 2136 bp in length
 * 45215 45314: gap of unknown length
 * 45315 47780: contig of 2466 bp in length
 * 47781 47880: gap of unknown length
 * 47881 50209: contig of 2329 bp in length
 * 50210 50309: gap of unknown length
 * 50310 52364: contig of 2055 bp in length
 * 52365 52464: gap of unknown length
 * 52465 54739: contig of 2275 bp in length
 * 54740 54839: gap of unknown length
 * 54840 56920: contig of 2081 bp in length
 * 56921 59676: contig of 2656 bp in length
 * 59677 59776: gap of unknown length
 * 59777 62058: contig of 2282 bp in length
 * 62059 62158: gap of unknown length
 * 62159 64481: contig of 2323 bp in length
 * 64482 64581: gap of unknown length
 * 64582 67737: contig of 3156 bp in length
 * 67738 67837: gap of unknown length
 * 67838 69856: contig of 2019 bp in length
 * 69857 69956: gap of unknown length
 * 69957 72698: contig of 2742 bp in length
 * 72699 72798: gap of unknown length
 * 72799 75900: contig of 3102 bp in length
 * 75901 76001: gap of unknown length
 * 76001 78614: contig of 2614 bp in length
 * 78615 78714: gap of unknown length
 * 78715 81699: contig of 2985 bp in length
 * 81700 81799: gap of unknown length
 * 81800 85910: contig of 4110 bp in length
 * 85910 86009: gap of unknown length
 * 86010 89796: contig of 3787 bp in length
 * 89797 89896: gap of unknown length
 * 89897 93635: contig of 3739 bp in length
 * 93636 97658: contig of 3923 bp in length
 * 97659 97758: gap of unknown length
 * 97759 101045: contig of 3287 bp in length

Scherer,S., Scott,G., Shen,H., Shoostari,N., Sisson,I.,
 Sodergren,E., Sonaik,T., Sparks,A., Stanley,H., Stone,H.,
 Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
 Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
 Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R., Wang,Q.,
 Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
 Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
 Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
 Weinstock,G. and Gibbs,R.
 Direct Submission
 Unpublished
 2 (bases 1 to 130899)
 Worley,K.C.
 Direct Submission
 Submitted (09-JUL-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 130899)
 Worley,K.C.
 Direct Submission
 Submitted (23-JUL-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Jul 17, 2002 this sequence version replaced gi:21716710.
 ----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GBWP
 Center clone name: CH230-6H4
 ----- Summary Statistics
 Sequencing vector: Plasmid;
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 71039 bases at least Q40
 Consensus quality: 77045 bases at least Q30
 Consensus quality: 80907 bases at least Q20

 * NOTE: Estimated insert size may differ from sequence length
 (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 69 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 * 1 1027: contig of 1027 bp in length
 * 1028 1127: gap of unknown length
 * 1128 2272: contig of 1145 bp in length
 * 2273 2372: gap of unknown length
 * 2373 3398: contig of 1026 bp in length
 * 3399 3498: gap of unknown length
 * 3499 4802: contig of 1304 bp in length
 * 4803 4902: gap of unknown length
 * 4903 5993: contig of 1091 bp in length
 * 5994 6093: gap of unknown length
 * 6094 7245: contig of 1151 bp in length
 * 7245 7345: gap of unknown length
 * 7345 8880: contig of 1536 bp in length
 * 8881 9881: gap of unknown length
 * 9881 9987: contig of 1007 bp in length
 * 9988 10087: gap of unknown length
 * 10088 11342: contig of 1255 bp in length
 * 11343 11442: gap of unknown length
 * 11443 12560: contig of 1118 bp in length
 * 12561 12660: gap of unknown length
 * 12661 13884: contig of 1224 bp in length
 * 13885 13984: gap of unknown length
 * 13985 15064: contig of 1080 bp in length

15065 15164: gap of unknown length
 15165 16546: contig of 1382 bp in length
 16547 16646: gap of unknown length
 16647 18133: contig of 1867 bp in length
 18134 18613: gap of unknown length
 18614 19705: contig of 1092 bp in length
 19706 19805: gap of unknown length
 19806 21270: contig of 1465 bp in length
 21271 21370: gap of unknown length
 21371 22600: contig of 1230 bp in length
 22601 22700: gap of unknown length
 22701 23741: contig of 1041 bp in length
 23742 23841: gap of unknown length
 23842 24945: contig of 1104 bp in length
 24946 25045: gap of unknown length
 25046 27358: contig of 2213 bp in length
 27359 28617: gap of unknown length
 28618 28717: contig of 1259 bp in length
 28718 28992: contig of 1175 bp in length
 28993 31396: gap of unknown length
 31397 31496: contig of 1404 bp in length
 31497 32504: contig of 1008 bp in length
 32505 32604: gap of unknown length
 32605 34001: contig of 1396 bp in length
 34001 35857: gap of unknown length
 35858 35957: contig of 1757 bp in length
 35958 37377: gap of unknown length
 37378 37477: contig of 1420 bp in length
 37478 38497: gap of unknown length
 38498 38597: contig of 1020 bp in length
 38598 38597: gap of unknown length
 38599 39793: contig of 1096 bp in length
 39794 42502: contig of 2709 bp in length
 42503 42602: gap of unknown length
 42603 44424: contig of 1821 bp in length
 44424 44523: gap of unknown length
 44524 46230: contig of 1707 bp in length
 46231 46330: gap of unknown length
 46331 47760: contig of 1429 bp in length
 47760 47860: gap of unknown length
 47860 49208: contig of 1349 bp in length
 49209 50360: gap of unknown length
 50360 50461: contig of 1052 bp in length
 50461 52257: gap of unknown length
 52258 52357: contig of 1797 bp in length
 52358 54147: gap of unknown length
 54148 54247: contig of 1790 bp in length
 54248 55402: gap of unknown length
 55403 55502: contig of 1155 bp in length
 55503 57008: gap of unknown length
 57009 57108: contig of 1506 bp in length
 57109 58547: gap of unknown length
 58548 58647: contig of 1439 bp in length
 58648 60159: gap of unknown length
 60160 60259: contig of 1512 bp in length
 60260 61728: gap of unknown length
 61729 61828: contig of 1469 bp in length
 61829 64395: gap of unknown length
 64396 64495: contig of 2567 bp in length
 64496 66096: gap of unknown length
 66097 66196: contig of 1601 bp in length
 66197 67712: gap of unknown length
 67713 67812: contig of 1516 bp in length
 67813 68997: gap of unknown length
 68998 69097: contig of 1185 bp in length
 69099 70867: gap of unknown length
 70868 70967: contig of 1770 bp in length
 70968 72129: gap of unknown length
 72129 72229: contig of 1162 bp in length
 72230 72310: gap of unknown length

WO9955882-A1.
04-NOV-1999.

```
PF 22-APR-1999; 99WO-US08791.
XX
XX 24-APR-1998; 98US-0082960.
XX
XX (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
XX Cahoon RE, Hitz WD, Thorpe CJ, Tingey SV;
XX
XX WPI; 2000-072181/06.
XX
XX P-PSDB; AAY44230.
XX
XX Novel phytic acid biosynthetic enzymes used to alter the level of the
XX enzyme in transgenic plants
XX
XX Claim 7; Page 37; 60pp; English.
XX
XX The present sequence is the cDNA insert in clone cdt2c.pk003.b20 encoding
XX corn extragenic suppressor protein. The cDNA clone was isolated from corn
XX developing tassels. The level of expression of this phytic acid
XX biosynthetic enzyme can be altered by transforming a host cell with the
XX chimeric gene comprising this cDNA. This is one to increase levels of
XX free phosphate and decrease phytic acid levels in developing seeds or in
XX the construction of transgenic plants with altered levels of the enzyme.
XX The phytic acid biosynthetic enzymes are used to raise antibodies, for
XX enzyme detection in cells and cell extracts. The polynucleotides also
XX serve as a source of probes and primers, for genetic mapping, as markers,
XX and to isolate homologous sequences from other species. This is
XX commercially used for animal feed.
XX
XX Sequence 492 BP; 88 A; 169 C; 144 G; 88 T; 3 other;

Query Match 99.4%; Score 489; DB 21; Length 492;
Best Local Similarity 100.0%; Pred. No. 9.7e-91;
Matches 492; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGCTCCGAGCTCATTTCTCCGCCCGACATTTAAACCTTGCTCCGACAAACCGCGCGA 60
Db 1 AGCTCCGAGCTCATTTCTCCGCCCGACATTTAAACCTTGCTCCGACAAACCGCGCGA 60
Qy 61 CTCCTCCCAATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 120
Db 61 CTCCTCCCAATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 120
Qy 121 CTTCCGCTCCGCAACCTTAACCCCGCTCTCGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 180
Db 121 CTTCCGCTCCGCAACCTTAACCCCGCTCTCGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 180
Qy 181 CTTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 240
Db 181 CTTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 240
Qy 241 GTTCCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 300
Db 241 GTTCCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 300
Qy 301 GTTGTGAGGTGGCGCAACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 360
Db 301 GTTGTGAGGTGGCGCAACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 360
Qy 361 CGCGCAGCGGGTTGAGATCATCGAAGAGAGACACAGTCTCTGTACAAATTCAGATAG 420
Db 361 CGCGCAGCGGGTTGAGATCATCGAAGAGAGACACAGTCTCTGTACAAATTCAGATAG 420
Qy 421 AAGAAGAGAGAGCAATGNGTCAGTTATCTAAGAGCTTCCCTACTCAAGCAATT 480
Db 421 AAGAAGAGAGAGCAATGNGTCAGTTATCTAAGAGCTTCCCTACTCAAGCAATT 480
Qy 481 TTGGNGAAGAGA 492
Db 481 TTGGNGAAGAGA 492

RESULT 2
```

```
AAZ29168
ID AAZ29168 standard; DNA; 1418 BP.
XX
XX AAZ29168;
XX
XX 21-FEB-2000 (first entry)
XX
XX Partial cDNA insert encoding wheat extragenic suppressor protein.
XX
XX Clone wlm0.pk0010.f6; wheat extragenic suppressor protein;
XX phytic acid biosynthetic enzyme; transformation; chimeric gene; antibody;
XX free phosphate; phytic acid level; transgenic plant; enzyme detection;
XX genetic mapping; animal feed; ds.
XX
XX Triticum aestivum.
XX
XX Key Location/Qualifiers
XX CDS 2..976
XX FT /tag= a
XX FT /product= "wheat extragenic suppressor protein"
XX FT /note= "Partial sequence"
XX
XX WO9955882-A1.
XX
XX 04-NOV-1999.
XX
XX 22-APR-1999; 99WO-US08791.
XX
XX 24-APR-1998; 98US-0082960.
XX
XX (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
XX Cahoon RE, Hitz WD, Thorpe CJ, Tingey SV;
XX
XX WPI; 2000-072181/06.
XX
XX P-PSDB; AAY44233.
XX
XX Novel phytic acid biosynthetic enzymes used to alter the level of the
XX enzyme in transgenic plants
XX
XX Claim 7; Page 40-41; 60pp; English.
XX
XX The present sequence is the cDNA insert in clone wlm0.pk0010.f6 encoding
XX a portion of wheat extragenic suppressor protein. The cDNA clone was
XX isolated from wheat seedlings. The level of expression of this phytic
XX acid biosynthetic enzyme can be altered by transforming a host cell with
XX the chimeric gene comprising this cDNA. This is done to increase levels
XX of free phosphate and decrease phytic acid levels in developing seeds or
XX in the construction of transgenic plants with altered levels of the
XX enzyme. The phytic acid biosynthetic enzymes are used to raise
XX antibodies, for enzyme detection in cells and cell extracts. The
XX polynucleotides also serve as a source of probes and primers, for genetic
XX mapping, as markers, and to isolate homologous sequences from other
XX species. This is commercially used for animal feed.
XX
XX Sequence 1418 BP; 393 A; 307 C; 386 G; 332 T; 0 other;

Query Match 33.7%; Score 165.8; DB 21; Length 1418;
Best Local Similarity 69.7%; Pred. No. 6.9e-25;
Matches 251; Conservative 0; Mismatches 105; Indels 4; Gaps 2;

Qy 136 CCCTAAACCCCGCTCTCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 195
Db 22 CCCTTACCACCTCGCTCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 81
Qy 196 CTTGAGCGGAGTGGGCGCCAGCCGATGAGTACGTTAGGGCCCTGTTTCGCCGCTGGGGC 255
Db 82 GGCTTGGGATAGCGGCGCCGTTGGATGGGCTCGGTTTCGAGCCCTCGCCTCTGAGGCGG 141
Qy 256 GGCGCGCGGAGAGCTGGCGCAGT---GGGGAGTTCGCGACGAGCGGCTGGTGGAGGT 312
Db 142 GGCTTGGGCGCTGGCTGGCGGGTAAGAGGGGGTGGACATGGAGCGGCTGGTGGCGGT 201
```

```
QY 313 GCGCAACGCGGCGGAGCGTCTGGGAGGTGCTCANCAAGTACTTCCCGCCAGCGGGT 372
||||| | ||||| | || ||||| ||||| ||||| ||||| | ||||| |
Db 202 GCGCGAGCGCGCGGATGCGCGCGGGGAGGTGCTCAGAAAGTACTTCAGCGACGCGTT 261
||||| | ||||| | || ||||| ||||| ||||| ||||| | ||||| |
QY 373 TGAGATCATCGACAAGAGGACCAAGTCTCTTTACAATTGAGATAGAGAAGCAGAAG 432
||||| | ||||| | || ||||| ||||| ||||| ||||| | ||||| |
Db 262 CGAGATCATCGACAAGAGGACCAAGTCTCTTTACAATTGAGATAGAGAAGCAGAAG 320
||||| | ||||| | || ||||| ||||| ||||| ||||| | ||||| |
QY 433 AAGCAATGNGTTCAGTTATCTACTGAAGAGCTTCCCTACTCTAA... 492
||||| | ||||| | || ||||| ||||| ||||| ||||| | ||||| |
Db 321 AAGCAATGACCTCAGTCTACTGTAAGAGCTTCCCTACTCTACTGCTCTTTTCGGTGAGGAGA 380

RESULT 3
AAZ29167
ID AAZ29167 standard; DNA; 1103 BP.
XX AC AAZ29167;
XX T 21-FEB-2000 (first entry)
XX A Assembled cDNA contig encoding soybean extragenic suppressor protein.
XX KW Clone s12.pk122.p24; clone src3c.pk013.g15; clone sfiln.pk003.q19;
XX KW soybean extragenic suppressor protein; phytic acid biosynthetic enzyme;
XX KW transformation; chimeric gene; antibody; free phosphate;
XX KW phytic acid level; transgenic plant; enzyme detection; genetic mapping;
XX KW animal feed; ds.
XX OS Glycine max.
XX FH Key Location/Qualifiers
XX FT 9..896
XX CDS /*tag= a
XX FT /product= "Soybean extragenic suppressor protein"
XX PN W09955882-A1.
XX XX 04-NOV-1999.
XX PD 22-APR-1999; 99WO-US08791.
XX PF 24-APR-1998; 98US-0082960.
XX PR (DUPO ) DU PONT DE NEMOURS & CO E I.
XX PA Cahoon RE, Hiltz WD, Thorpe CJ, Tingey SV;
XX PI WPI: 2000-072181/06.
XX JR P-PSDB; AAY44232.
XX XX Novel phytic acid biosynthetic enzymes used to alter the level of the
XX FT enzyme in transgenic plants -
XX PS Claim 7; Page 39; 60pp; English.
XX CC The present sequence is a contig assembled from portions of the cDNA
XX CC inserts in clones s12.pk122.p24, src3c.pk013.g15, and sfiln.pk003.q19,
XX CC encoding a soybean extragenic suppressor protein. The level of expression
XX CC of this phytic acid biosynthetic enzyme can be altered by transforming a
XX CC host cell with the chimeric gene comprising this cDNA. This is done to
XX CC increase levels of free phosphate and decrease phytic acid levels in
XX CC developing seeds or in the construction of transgenic plants with altered
XX CC levels of the enzyme. The phytic acid biosynthetic enzymes are used to
XX CC raise antibodies, for enzyme detection in cells and cell extracts, the
XX CC polynucleotides also serve as a source of probes and primers, for genetic
XX CC mapping, as markers, and to isolate homologous sequences from other
XX CC species. This is commercially used for animal feed.
XX SQ Sequence 1103 BP; 307 A; 238 C; 236 G; 322 T; 0 other;
Query Match 14.0%; Score 69; DB 21; Length 1103;
Best Local Similarity 62.2%; Pred. No. 3.4e-05;
```

```
Matches 122; Conservative 0; Mismatches 73; Indels 1; Gaps 1;
QY 297 AGCGGCTGTTGGAGTGGCGAACGGGGCGGAGCGTCTGGGAGGTGCTCANGAAGT 356
||||| | ||||| | || ||||| ||||| ||||| ||||| | ||||| |
Db 106 ATCGCTTCGCGAGTTCGTTACAAAGTCGCCGATGCTGCGGAGAGTTTATCCGCAAT 165
||||| | ||||| | || ||||| ||||| ||||| ||||| | ||||| |
QY 357 ACTTCCGCCAGCGGGTTGAGATCATCGACAAGAGGACCAAGTCTCTGTTCATAATTGCAG 416
||||| | ||||| | || ||||| ||||| ||||| ||||| | ||||| |
Db 166 ACTTCAGAAAAAACTTCGACGTTATTCAAAACATGATCTCAGTCCAGTAACCATTCGAG 225
||||| | ||||| | || ||||| ||||| ||||| ||||| | ||||| |
QY 417 ATAGAAGAAGCAGAGAAGCAATGNGTCAAGTCTTATCTAGAGAGCTTCCCTACTCAAGCC 476
||||| | ||||| | || ||||| ||||| ||||| ||||| | ||||| |
Db 226 AT-CAATCTGCTGAGGAGGCTATGTTCAATCATAGACAAATTTCCCTTCTCATGCC 284
||||| | ||||| | || ||||| ||||| ||||| ||||| | ||||| |
QY 477 AATTTTGGNGAAGAGA 492
||||| | ||||| | || ||||| ||||| ||||| ||||| | ||||| |
Db 285 ATTTACGAGAGGAGAA 300

RESULT 4
AAV62176
ID AAV62176 standard; DNA; 117213 BP.
XX AC AAV62176;
XX XX 13-JAN-1999 (first entry)
XX DE HSV-2 strain SB5 Contig ID 15 DNA sequence.
XX KW HSV-2 strain SB5; immunological response induction; therapy;
XX KW antiviral identification; viral protein inhibitor; ss.
XX OS Herpes simplex virus type 2.
XX FH Key Location/Qualifiers
XX FT 755..1297
XX CDS /*tag= a
XX FT /product= "ORF#1 protein"
XX FT /note= "encoded protein shown in AAW72170"
XX FT 1170..2174
XX CDS /*tag= b
XX FT /product= "ORF#2 protein"
XX FT /note= "encoded protein shown in AAW72171"
XX FT 2229..2930
XX FT /*tag= c
XX FT /product= "ORF#3 protein"
XX FT /note= "encoded protein shown in AAW72172"
XX FT complement (4130..3735)
XX FT /*tag= d
XX FT /product= "ORF#4 protein"
XX FT /note= "encoded protein shown in AAW72173"
XX FT complement (3802..6447)
XX FT /*tag= e
XX FT /product= "ORF#5 protein"
XX FT /note= "encoded protein shown in AAW72174"
XX FT 6017..8482
XX CDS /*tag= f
XX FT /product= "ORF#6g protein"
XX FT /note= "encoded protein shown in AAW72250"
XX FT 6026..8482
XX FT /*tag= g
XX FT /product= "ORF#6f protein"
XX FT /note= "encoded protein shown in AAW72249"
XX FT 6065..8482
XX CDS /*tag= h
XX FT /product= "ORF#6e protein"
XX FT /note= "encoded protein shown in AAW72248"
XX FT 6167..8482
XX CDS /*tag= i
XX FT /product= "ORF#6d protein"
XX FT /note= "encoded protein shown in AAW72247"
XX FT 6296..8482
XX CDS /*tag= j
```

```
FT FT /product= "ORF#6c protein" CDS FT FT complement (27630..31784)
FT FT /note= "encoded protein shown in AAW72246" FT FT /*tag= ab
FT FT 6326..8482 FT FT /product= "ORF#20a protein" FT FT /note= "encoded protein shown in AAW72224"
FT FT /*tag= k FT FT complement (32067..32735)
FT FT /product= "ORF#6b protein" CDS FT FT /*tag= ac
FT FT 6446..8482 FT FT /product= "ORF#21 protein" FT FT /note= "encoded protein shown in AAW72187"
FT FT /*tag= i FT FT 33140..34984
FT FT /product= "ORF#6a protein" CDS FT FT /*tag= ad
FT FT /transl_except= (pos: 7400..7402, aa: Ala-Ala) FT FT /product= "ORF#22a protein" FT FT /note= "encoded protein shown in AAW72226"
FT FT /transl_except= (pos: 7481..7486, aa: Ile) FT FT 33386..34984
FT FT /note= "encoded protein shown in AAW72244" CDS FT FT /*tag= ae
FT FT 8457..9347 FT FT /product= "ORF#22b protein" FT FT /note= "encoded protein shown in AAW72227"
FT FT /*tag= m FT FT complement (9604..11855)
FT FT /product= "ORF#7 protein" CDS FT FT /*tag= n
FT FT /note= "encoded protein shown in AAW72175" FT FT /transl_except= (pos: 11635..11636, aa: Ala)
FT FT complement (9604..11855) FT FT /product= "ORF#8 protein" FT FT /note= "encoded protein shown in AAW72176"
FT FT /*tag= o FT FT complement (11905..14508)
FT FT /product= "ORF#9a protein" CDS FT FT /*tag= p
FT FT /note= "encoded protein shown in AAW72223" FT FT /product= "ORF#9b protein" FT FT /note= "encoded protein shown in AAW72222"
FT FT 14399..15802 FT FT /*tag= q FT FT complement (11905..14520)
FT FT /product= "ORF#10 protein" CDS FT FT /*tag= r
FT FT /note= "encoded protein shown in AAW72177" FT FT /product= "ORF#9a protein" FT FT /note= "encoded protein shown in AAW72223"
FT FT complement (15996..16286) FT FT /*tag= s FT FT 14399..15802
FT FT /product= "ORF#11 protein" CDS FT FT /*tag= t
FT FT /note= "encoded protein shown in AAW72178" FT FT complement (16202..18064)
FT FT /product= "ORF#12 protein" CDS FT FT /*tag= u
FT FT /note= "encoded protein shown in AAW72179" FT FT /product= "ORF#13 protein" FT FT /note= "encoded protein shown in AAW72180"
FT FT complement (18105..19661) FT FT /*tag= v FT FT 20155..21453
FT FT /product= "ORF#14 protein" CDS FT FT /*tag= w
FT FT /note= "encoded protein shown in AAW72181" FT FT /product= "ORF#15 protein" FT FT /note= "encoded protein shown in AAW72182"
FT FT 20155..21453 FT FT complement (21326..22291)
FT FT /product= "ORF#16 protein" CDS FT FT /*tag= x
FT FT /note= "encoded protein shown in AAW72183" FT FT complement (22546..24654)
FT FT /product= "ORF#17 protein" CDS FT FT /*tag= y
FT FT /note= "encoded protein shown in AAW72184" FT FT /product= "ORF#18 protein" FT FT /note= "encoded protein shown in AAW72185"
FT FT 24684..25955 FT FT /*tag= z
FT FT /product= "ORF#19 protein" CDS FT FT /note= "encoded protein shown in AAW72186"
FT FT complement (26295..27251) FT FT /*tag= aa
FT FT /product= "ORF#20b protein" FT FT /note= "encoded protein shown in AAW72225"
FT FT 27630..31754
```


PN W0905272-A1.
XX 04-FEB-1999.
XX 27-JUL-1998; 98WO-GB02228.
XX 10-DEC-1997; 97US-0988476.
PR 25-JUL-1997; 97GB-0015823.
XX (UNLO) UNIV COLLEGE LONDON.
XX Latchman DS, Smith MD;
XX WPI; 1999-142928/12.
DR P-PSDB; AAW96263.
XX
XX New polypeptide comprising transcription factor Brn-3a, or its
PT derivative - useful for treating nervous system diseases, preventing
PT cellular apoptosis and increasing nerve regeneration following
PT neuronal damage
XX
XX Disclosure; Page 63-64; 68pp; English.
XX
XX Over expression of transcription factor Brn-3a can protect neuronal
CC cells from apoptosis. Brn-3a also specifically activates expression
CC of the Bcl-2 gene in neuronal cells and this activation is mediated
CC via a Brn-3a response element in the 5' regulatory region of the
CC Bcl-2 gene. Both the anti-apoptotic effect of Brn-3a and its ability
CC to activate expression of Bcl-2 are mediated by the N-terminal
CC domain of Brn-3a. Members of the Bcl-2 family perform critical roles
CC in the regulation of selective apoptosis during development of the
CC nervous system. The stimulation of Bcl-2 expression by Brn-3a in a
CC neuron specific manner and consequent protection of neuronal cells
CC from apoptosis suggests that Brn-3a may co-ordinate some aspects of
CC neuronal reorganisation during development or following injury. The
CC elevation of Brn-3a expression by either pharmacological means
CC (compositions comprising one other therapeutic polypeptide e.g.
CC neurotrophic factors, nerve growth factor (NGF), ciliary neurotrophic
CC factor (CNTF), brain-derived neurotrophic factor (BDNF), and
CC neurotrophins NT-3 and NT-4/5) or gene therapy may represent a
CC method for treating human diseases associated with excessive
CC neuronal cell death and/or lack of nerve regeneration, especially
CC neurodegenerative diseases such as familial dysautonomia and
CC infantile muscular dystrophy, and Parkinson's and Alzheimer's
CC disease.
XX
XX Sequence 1266 BP; 201 A; 455 C; 458 G; 152 T; 0 other;
XX
XX Query Match 11.8%; Score 58.2; DB 20; Length 1266;
XX Best Local Similarity 51.3%; Pred. No. 0.0055;
XX Matches 135; Conservative 0; Mismatches 128; Indels 0; Gaps 0;
XX
QY 40 GCTCCGACACCGCGCGACTCTCCGCCAATGCTCTCTCTCTCTCCACCACTC 99
DB 550 GCCCG 491
QY 100 GGCCACTCGCCCTTCCCGGCGCTCGCTTCGCCAACCCTAACCCCGCTCTCGCTCT 159
DB 490 GGCG 431
QY 160 CCGCTCCCG 219
DB 430 GGCG 371
QY 220 GATGAGTAGGGTTAGGGCTCTTCCCGCTTGGGGCGCGCGCGAGAGCTGCGGCGAGT 279
DB 370 ACGGCCACGAGATGTGTCCAGCAGTACCGGCTCGAGCGCTTGGTGGTGGTGTGT 311
QY 280 GGGGAGTTGGGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 302
DB 310 GGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 288

RESULT 9
AAA29007/c
ID AAA29007 standard; cDNA; 1266 BP.
XX
XX AC AAA29007;
XX
XX 12-SEP-2000 (first entry)
XX Murine transcription factor Brn-3a coding sequence.
XX
XX Brn-3a; modulator; inhibitor; cervical cancer; human papilloma virus;
KW HPV; antisense; cytostatic; ss.
XX
XX Mus musculus.
XX
XX Key Location/Qualifiers
FT CDS 1..1266
FT /tag= a
FT /product= Brn-3a_transcription_factor
XX
XX W0200034466-A1.
XX
XX 15-JUN-2000.
XX
XX 07-DEC-1999; 99WO-GB04116.
XX
XX 07-DEC-1998; 98GB-0026888.
PR 31-MAR-1999; 99US-0282210.
XX
XX (UNLO) UNIV COLLEGE LONDON.
XX
XX Latchman DS, Budhram-Mahadeo V, Ndisang D;
XX WPI; 2000-423416/36.
DR P-PSDB; AAY96405.
XX
XX Product for treating, preventing and diagnosing cervical cancer
PT comprises a nucleotide sequence or molecule which binds to Brn-3a,
PT decreases its intracellular levels or inhibits its activity
XX
XX Disclosure; Page 65-68; 72pp; English.
XX
XX A product that binds, causes a decrease in intracellular levels of or
CC inhibits the activity of Brn-3a useful for treating, prevention or
CC diagnosis of cervical cancer caused by human papilloma virus (HPV) is
CC claimed. Expression of HPV proteins is generally dependent on the
CC presence of Brn-3a in the cell. Methods of identifying Brn-3a binding
CC agents or agents which inhibit Brn-3a expression are claimed. Nude mice
CC were injected with SiHa cells containing a single integrated HPV16-genome
CC were transformed with a Brn-3a antisense construct and with the empty
CC expression vector as control and tumours assessed at regular intervals.
CC Results showed that after 30 days there was no or very little tumour
CC growth in mice transformed with Brn-3a antisense construct as compared
CC to the control.
XX
XX Sequence 1266 BP; 201 A; 455 C; 458 G; 152 T; 0 other;
XX
XX Query Match 11.8%; Score 58.2; DB 21; Length 1266;
XX Best Local Similarity 51.3%; Pred. No. 0.0055;
XX Matches 135; Conservative 0; Mismatches 128; Indels 0; Gaps 0;
XX
QY 40 GCTCCGACACCGCGCGACTCTCCGCCAATGCTCTCTCTCTCTCCACCACTC 99
DB 550 GCCCG 491
QY 100 GGCCACTCGCCCTTCCCGGCGCTCGCTTCGCCAACCCTAACCCCGCTCTCGCTCT 159
DB 490 GGCG 431
QY 160 CCGCTCCCG 219
DB 430 GGCG 371

```
QY 220 GATGAGTACGTTAGCGCTTCGCTGGCGCGCGGAGACTGCGGCAGT 279
Db 370 ACGCCGACGAGATGTCTCAGCAGGTACCGGGCTCGACCGCTGTGTGTGGTGGT 311
QY 280 GGGGAGTTGGCGACGAGCGGC 302
Db 310 GGTGTGTGTGTGGCGCGGC 288

RESULT 10
ID ABK42449
XX ABK42449 standard; DNA; 9953 BP.
AC ABK42449;
XX
XX
DT 21-MAY-2002 (first entry)
DE
DE Genomic sequence #348 encoding novel human connective tissue polypeptide.
XX
XX Human; connective tissue related disorder; cancer; gene therapy;
XX cytosstatic; gene; ds.
XX
OS Homo sapiens.
XX
XX WO200155343-A1.
PN
XX
PD 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US01322.
XX
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-02 063.
PR 14-SEP-2000; 2000US-02 064.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236377.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239335.
PR 13-OCT-2000; 2000US-0239337.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0244674.
PR 08-NOV-2000; 2000US-024475.
PR 08-NOV-2000; 2000US-024476.
PR 08-NOV-2000; 2000US-024477.
PR 08-NOV-2000; 2000US-024478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
```

PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
X (HUMA-) HUMAN GENOME SCI INC.
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-565190/63.
DR Nucleic acid encoding novel connective tissue associated polypeptides,
XX used in diagnosing, preventing, treating or ameliorating a disorder
PT such as cancer or rheumatoid arthritis -
PT Disclosure: SEQ ID No 1336; 673pp; English.
PS The present invention relates to the isolation of novel human connective
XX tissue related polypeptides (AAU86435-AAU86423) and the polynucleotide
CC (cDNA and genomic) sequences encoding them. The sequences of the
CC invention are useful in the diagnosis, treatment, prevention and/or
CC prognosis of diseases associated with connective tissue(s), including
CC cancer. The polynucleotide sequences of the invention are also useful
CC in gene therapy. ABK42102-ABK43116 represent genomic sequences encoding
CC the novel human connective tissue related polypeptides.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pt_sequences.
SQ Sequence 9953 BP; 1958 A; 3052 C; 2783 G; 2160 T; 0 other;
Query Match 11.5%; Score 56.4; DB 23; Length 9953;
Best Local Similarity 58.2%; Pred. No. 0.016;
Matches 99; Conservative 0; Mismatches 71; Indels 0; Gaps 0;
QY 21 GCCCGGACATTTAAACCTTGTCGCCGACACCGCGCGGACTGCTCCCAATGCTCTCCT 80
Db 3991 GCTTTAGCTTGGAGGCTTCTGCGCGACGCGCGGCTCTCTCTCTCTCTCTCTCTCT 4050
QY 81 CTCTCTCTCTCCACCTGCTCCGCGACCTGCTCTCCGCGGCTGCTCCGCAACCCCTA 140
Db 4051 CT 4110
QY 141 ACCCCGCGCTGCTGCT 190
Db 4111 CCT 4160
RESULT 11
ABK42448
ID ABK42448 standard; DNA; 11189 BP.
XX AC
XX ABK42448;
XX AC
XX DT 21-MAY-2002 (first entry)
XX DE Genomic sequence #347 encoding novel human connective tissue polypeptide.

XX Human; connective tissue related disorder; cancer; gene therapy;
KW cytostatic; gene; ds.
KW Homo sapiens.
XX WO200155343-A1.
XX 02-AUG-2001.
XX 17-JAN-2001; 2001WO-US01322.
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184564.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0227182.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 05-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.

PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.

PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-565190/63.
XX Nucleic acid encoding novel connective tissue associated polypeptides,
XX used in diagnosing, preventing, treating or ameliorating a disorder
XX such as cancer or rheumatoid arthritis -
XX Disclosure; SEQ ID NO 1335; 673pp; English.
XX The present invention relates to the isolation of novel human connective
XX tissue related polypeptides (AAU86435-AAU86923) and the polynucleotide
XX (cDNA and genomic) sequences encoding them. The sequences of the
XX invention are useful in the diagnosis, treatment, prevention and/or
XX prognosis of diseases associated with connective tissue(s), including
XX cancer. The polynucleotide sequences of the invention are also useful
XX in gene therapy. ABK42102-ABK43116 represent genomic sequences encoding
XX the novel human connective tissue related polypeptides.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 11189 BP; 2196 A; 3373 C; 3188 G; 2432 T; 0 other;
Query Match 11.5%; Score 56.4; DB 23; Length 11189;
Best Local Similarity 58.2%; Pred. No. 0.016;
Matches 99 Conservative 0; Mismatches 71; Indels 0; Gaps 0;
QY 21 GCCCGACATTAAACCTTCCTCCGACAAACGGCGGCGACTCTCCCAATGCTCTCCT 80
DB 5227 GCTTGGAGCTGGAGGCTTCCTGCCCGACGCGCGGCTCTCTCTCTCTCTCTCT 80
QY 81 CTTCTCTCTCCACCCACTCGGCGACCTCGCCCTTCGCCGCGCTCGGCTCGGCAACCCCTA 140
DB 5287 CTTCT 140
QY 141 ACCCGCGCTCTCGCTCTCTCGGCTCGGCGCGGCGGCTCGGCGGCTCTCTCTCTCTCT 190
DB 5347 CTTCT 5396
RESULT 12
AA184343/C
ID AA184343 standard; cDNA; 434 BP.
XX
XX AA184343;
XX
XX 06-NOV-2001 (first entry)
XX Human polynucleotide SEQ ID NO 4403.
XX
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
XX vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
XX tissue growth factor; immunomodulatory; cancer; leukaemia;
XX nervous system disorders; arthritis; inflammation; ss.
XX Homo sapiens.
XX
XX WO200164835-A2.
XX
XX 07-SEP-2001.
XX
XX 26-FEB-2001; 2001WO-US04927.
XX
XX 28-FEB-2000; 2000US-0515126.
XX 18-MAY-2000; 2000US-0577409.
XX
XX (HYSE-) HYSEQ INC.

```
XX Tang YT, Liu C, Drmanac RT;
XX WPI; 2001-514838/56.
XX P-PSDB; AAO04412.
XX Isolated nucleic acids and polypeptides, useful for preventing
XX diagnosing and treating e.g. leukaemia, inflammation and immune
XX disorders.
XX Claim 1; SEQ ID NO 4403; 1399pp + Sequence Listing; English.
XX The invention relates to human polynucleotides (AAI79941-AA193841) and
XX the encoded proteins (AAO0010-AAO13910) that exhibit activity relating to
XX cytokine, cell proliferation or cell differentiation or which may induce
XX production of other cytokines in other cell populations. The
XX polynucleotides and polypeptides are useful in gene therapy, vaccines or
XX peptide therapy. The polypeptides have various cytokine-like activities,
XX e.g. stem cell growth factor activity, haematopoiesis regulating
XX activity, tissue growth factor activity, immunomodulatory activity and/or
XX activin/inhibin activity and may be useful in the diagnosis and/or
XX treatment of cancer, leukaemia, nervous system disorders, arthritis and
XX inflammation.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX Sequence 434 BP; 34 A; 32 C; 256 G; 47 T; 65 other;
XX
XX Query Match 11.4%; Score 56; DB 22; Length 434;
XX Best Local Similarity 41.8%; Pred. No. 0.014;
XX Matches 107; Conservative 0; Mismatches 149; Indels 0; Gaps 0;
XX
QY 17 CTCGCCCGGACATTTAAACCTTCTCCGACACCGCCCGGACTCTCTCCCAATGCTC 76
DB 11 ||||| || ||||| || ||||| || ||||| || ||||| ||
QY 387 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 328
DB CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
QY 77 TCTCTTCTCTCTACCCACTCGGCGACCTCGGCTTCCCGGCTTCGCTCGGAAAC 136
DB || ||||| || ||||| || ||||| || ||||| || ||||| ||
QY 327 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 268
DB CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
QY 137 CCTAACCCCGGCTCTCGCTCTCTCGGCTCGGCGCCGCTCGCGGTGTGTGTCGCGGTC 196
DB || ||||| || ||||| || ||||| || ||||| || ||||| ||
QY 267 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 208
DB CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
QY 197 TTGAGCGGAGTGGCGCGACGCGATGAGTACGGTTAGGGCTCGTTTCGCGCGTGGGCG 256
DB || ||||| || ||||| || ||||| || ||||| || ||||| ||
QY 257 GCCGGCGGAGAGCTG 272
DB 147 CCCCACCCCCAGCTG 132
XX
RESULT 13
AB199656/c
ID AB199656 standard; cDNA; 1286 BP.
XX
XX AB199656;
XX
XX 07-MAR-2002 (first entry)
XX
XX Mouse ischaemic condition related cDNA sequence SEQ ID NO:686.
XX
XX Mouse; ischaemia; compressive ischaemia; occlusive ischaemia;
XX vasospastic ischaemia; ischaemic condition; ischaemic disease; ss.
XX
XX Mus musculus.
XX
XX WO200188188-A2.
XX
XX 22-NOV-2001.
XX
```

```
PF 18-MAY-2001; 2001WO-JP04192.
PR 18-MAY-2000; 2000JP-0145977.
PA (UYNI-) UNIV NIHON SCHOOL JURIDICAL PERSON.
PI Ishikawa K, Asai S, Takahashi Y, Nagata T, Ishii Y;
XX WPI; 2002-034733/04.
XX Examining the ischemic condition (e.g. occlusive ischemia) by measuring
XX expression levels of particular genes defined in the specification or
XX by determining the expression profile of a gene group comprising these
XX genes.
XX Claim 2; Page 1713-1714; 2690pp; English.
XX The present invention describes a method for examining ischaemic
XX conditions, comprising measuring the expression levels of particular
XX genes (I) in a test sample or determining the expression profile of a
XX gene group in the sample comprising genes selected from (I). The method
XX is useful for examining the ischaemic condition (e.g. compressive
XX ischaemia, occlusive ischaemia or vasospastic ischaemia) by measuring
XX expression levels of particular genes (ABI99202 to ABI9912, encoding
XX the protein sequences in ABB57020 to ABB57374) or by determining the
XX expression profile of a gene group comprising these genes. The
XX expression levels or expression profiles produced by these genes are
XX used as an indicator when screening for ischaemic condition-improving
XX drugs or therapeutics for ischaemic diseases. ABI9913 and ABI9914
XX represent PCR primers for a mouse ischaemic condition related sequence,
XX which are used in the exemplification of the present invention.
XX Sequence 1286 BP; 239 A; 252 C; 462 G; 173 T; 160 other;
XX
XX Query Match 11.4%; Score 56; DB 24; Length 1286;
XX Best Local Similarity 32.4%; Pred. No. 0.016;
XX Matches 110; Conservative 0; Mismatches 229; Indels 0; Gaps 0;
XX
QY 3 CTCGAGCGTCACTTCCTCCGCGGACATTTAAACCTTCTCCGCGGCGGCGGCT 62
DB 1257 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 1198
QY 63 CTTCCCAATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 122
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
QY 1197 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 1138
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
QY 123 TCGCTCCGCAACCCCTAATCCGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 182
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
QY 1137 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 1078
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
QY 183 TGTCTCTCCGCGGCTTTGAGCGGAGTGGGCGGCGGAGTACGATGAGTGGGCTCGT 242
DB 1077 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 1018
QY 243 TCGCGCTGGGCGGCGGCGGCGGAGAGCTCGGCGAGTGGGCGGAGTGGGCGGCG 302
DB 1017 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 958
QY 303 TGTGGAGTGGGCGCAACGCGGCGGCGGAGCTGCTGCTGGG 341
DB 957 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNGTG 919
XX
RESULT 14
AAS67813
ID AAS67813 standard; cDNA; 853 BP.
XX
XX AAS67813;
XX
XX 13-FEB-2002 (first entry)
XX
XX DNA encoding novel human diagnostic protein #3617.
XX
```


	Qy	Db	Qy	Db
140	AACCCCGCTCTGCCTCTCCGCTCCGCGCGCTGCTGCGCGGTCGTG	199	AGCGCGAGTGGGCGCCAGCCGATGAGTA	227
924	CC	865	TCCGTGGGCGCGCTGAAGTTCCTG	837

Search completed: November 19, 2002, 00:26:51
Job time : 396 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 17, 2002, 14:30:07 ; Search time 26 Seconds
(without alignments)
539.192 Million cell updates/sec

Title: US-09-686-522c-14

Perfect score: 1730

Sequence: 1 M5SSSTHSDTSPFGLAS.....VVAAGDARVHKALDALRWR 338

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen paramete: 3; 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	334	19.3	261	1 YHEB_CHLVI	P56160 Chlorobium
2	292	16.9	271	1 SUHB_PSEAE	Q9hxi4 pseudomonas
3	287	16.6	267	1 SUHB_PASMU	Q9cnv8 pasteurrella
4	279	16.1	264	1 SUHB_AQUAE	O67791 aquifex ao
5	272	15.7	267	1 SUHB_HAEIN	P44333 haemophilus
6	265.5	15.3	268	1 MYO3_LYCES	P54928 lycopersico
7	262	15.1	267	1 SUHB_SALTY	P58537 salmonella
8	258	14.9	267	1 SUHB_ECOLI	P22783 escherichia
9	248	14.3	267	1 SUHB_VIBCH	Q9kty5 vibrio chol
10	247.5	14.3	273	1 MYO1_LYCES	P54926 lycopersico
11	238	13.8	265	1 MYO2_LYCES	P54927 lycopersico
12	235	13.6	270	1 MYOP_MESCR	O49071 mesembryant
13	234	13.5	275	1 SUHB_XYLFA	Q9pam0 xyliella fas
14	229.5	13.3	256	1 SUHB_THEMEA	O33832 thermotoga
15	228	13.2	266	1 SUHB_RHILO	Q98f59 rhizobium l
16	221.5	12.8	265	1 SUHB_BACSU	Q45499 bacillus su
17	218	12.6	259	1 STRO_STRGR	P29785 streptomyc
18	212.5	12.3	287	1 SUHB_SNVY3	P74158 synecocyst
19	211	12.2	266	1 SUHB_RHIME	Q92m71 rhizobium m
20	211	12.2	277	1 MYOP_BOVIN	P20456 bos taurus
21	209	12.1	285	1 MYOP_XENLA	P29219 xenopus lae
22	206	11.9	246	1 CYSQ_ECOLI	P22255 escherichia
23	203.5	11.8	277	1 MYOP_MOUSE	O55023 mus musculus
24	202.5	11.7	261	1 SUHB_NEIMB	Q91z07 neisseria m
25	201.5	11.6	261	1 SUHB_NEIMA	Q91z03 neisseria m
26	200	11.6	252	1 SUHB_METJA	Q57573 methanococc
27	196.5	11.4	288	1 MYO2_HUMAN	O14732 homo sapien
28	193.5	11.2	277	1 MYOP_HUMAN	P29218 homo sapien
29	193.5	11.2	277	1 MYOP_RAT	P97697 rattus norv
30	190.5	11.0	269	1 CYSQ_ACTAC	P70714 actinobacil
31	190	11.0	246	1 CYSQ_SALTY	P26264 salmonella
32	183.5	10.6	256	1 SUHB_CAUCR	Q9a3d5 caulobacter
33	181.5	10.5	341	1 YSNW_CAEEL	Q19420 caenorhabdi

ALIGNMENTS

RESULT 1

YHEB_CHLVI

ID	YHEB_CHLVI	STANDARD	PRT	261 AA.
AC	P56160;			
DT	01-NOV-1997 (Rel. 35, Created)			
DT	01-NOV-1997 (Rel. 35, Last sequence update)			
DT	15-DEC-1998 (Rel. 37, Last annotation update)			
DE	Hypothetical 28.2 kDa protein in hemb 3'/region.			
OS	Chlorobium vibrioforme.			
OC	Bacteria; Chlorobi; Chlorobia; Chlorobiales; Chlorobiaceae; Chlorobium.			
OX	NCBI_TaxID=1098;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=f thiosulfatophilum / NCIB 8327;			
RX	MEDLINE=J6215215; PubMed=8626508;			
RA	Rhie G.-E., Avissar Y.J., Beale S.I.;			
RT	*Structure and expression of the Chlorobium vibrioforme hemb gene and characterization of its encoded enzyme, porphobilinogen synthase.*;			
RL	J. Biol. Chem. 271:8176-8182(1996).			
CC	-f- SIMILARITY: BELONGS TO THE INOSITOL MONOPHOSPHATASE FAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; U38348; ; NOT_ANNOTATED_CDS.			
DR	HSSP; P29218; 1IMP.			
DR	InterPro; IPR000760; Inositol_P.			
DR	Pfam; PF00459; inositol_P; 1.			
DR	PROSITE; PS00629; IMP_1; 1.			
DR	PROSITE; PS00630; IMP_2; 1.			
KW	Hypothetical protein.			
SQ	SEQUENCE 261 AA; 28243 MW; D25E7D223024095F CRC64;			

Query Match

Best Local Similarity 19.3%; Score 334; DB 1; Length 261;

Matches 78; Conservative 43; Mismatches 91; Indels 18; Gaps 4;

QY	81	VAGRAADAAGEVLRYKF-RORVEIDKEDSPVTIADREAEAMWSVILKSFTHAIFGE	139
DB	9	LAELEAKAKGLTLDYFGRSLQVFSKRDDTPYTEADRAEELIRQGISAKFPDDGLFGE	68
QY	140	E-----NGWRCAENSADFVWVLDPIDGTSKPSITCKPLFGTLIALHNGKPVIGIDQP	192
DB	69	EFEDEHPSGNGRR-----WIIDPIDGTSRPHGVPLVGMIALEVEGAMQLGVINFP	119
QY	193	ILRRRIGVGDGKQTLINGOEISVRS-CNLLAQAYLTTPPHLEAEDAFIVRNKVKY	251
DB	120	ALGELQAEGRSGAFWNGSPVQVSAEANSASTVTEKEYTEKLLDPPSNIPVDQLDAGL	179
QY	252	PLYGDCYAYALLASGFVDIVVESGLKPYDFLSLVPVIEGAGGSITDWRG	301

RESULT 5
SUHB_HAEIN

```
Db 232 -----AGNSYLSRGNIIAA-PSRVIKEMLNKIR 258
RESULT 6
MYO3_LYCES
ID MYO3_LYCES STANDARD; PRT; 268 AA.
AC P54928;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myo-inositol-1(or 4)-monophosphatase 3 (EC 3.1.3.25) (IMPase 3) (IMP
DE 3) (inositol monophosphatase 3).
GN IMP3.
OS Lycopersicon esculentum (Tomato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
OX NCBI_TaxID=4081;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. VFNT Cherry;
RC MEDLINE=96351935; PubMed=8718627;
RA Gillaspay G.E., Keddie J.S., Oda K., Gruissem W.;
RT "Plant inositol monophosphatase is a lithium-sensitive enzyme encoded
RT by a multigene family.";
RL Plant Cell 7:2175-2185(1995).
CC -!- FUNCTION: IT IS RESPONSIBLE FOR THE PROVISION OF INOSITOL REQUIRED
CC FOR SYNTHESIS OF PHOSPHATIDYLINOSITOL AND POLYPHOSPHOINOSITIDES.
CC -!- CATALYTIC ACTIVITY: Myo-inositol 1-monophosphate + H(2)O -> myo-
CC inositol + phosphate.
CC -!- COFACTOR: REQUIRES MAGNESIUM (BY SIMILARITY).
CC -!- ENZYME REGULATION: INHIBITED BY Li(+).
CC -!- PATHWAY: KEY ENZYME OF THE PHOSPHATIDYL INOSITOL SIGNALING
CC PATHWAY.
CC -!- SIMILARITY: BELONGS TO THE INOSITOL MONOPHOSPHATASE FAMILY.
CC
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U39059; AAB19031.1; -.
CC HSP: P29218; 1MF.
CC InterPro: IPR000760; Inositol_P.
CC Pfam: PF00459; inositol_P; 1.
CC PROSITE: PS00629; IMP_1; 1.
CC PROSITE: PS00630; IMP_2; 1.
CC HydroLase; Magnesium; Lithium; Multigene family.
CC KW SEQUENCE 268 AA; 29054 MW; BCB635A029804901 CRC64;
Query Match 15.3%; Score 265.5; DB 1; Length 268;
Best Local Similarity 30.2%; Pred. No. 1.9e-13;
Matches 83; Conservative 48; Mismatches 117; Indels 27; Gaps 8;
QY 74 ATERLVEAQAADAAGEVLKRYFRQVEIIDKEDHSPVTIADREAEAMVSVILKSFFT 133
DB 6 SVEQFLQVAVAEAKKAGEIIRREGYKTKHVEHKGWDLVTETDKACEDFIHNLKQRFPS 65
QY 134 HAIFGEENGWRC--ENSADFVWVLDPIDGTSFKTFGLFGLIALLHNGKPVIGVIDQ 191
DB 66 HKFTEETAAACGNFELTDEPTWIDPLDGTTFVHGFPFVCVSGILTEKKPTGVVYN 125
QY 192 PILRERWIGVDKOTTLNGOEISVRSNLLAQAYLYT---TSPHLFEA-----EDAFI 243
DB 126 PIDLEFTGIDGKAFNGKPIKVSSELSVKKALLATEAGTNRDKLVVDATTGRINSLLF 185
QY 244 RVRNKKVPLVYGCDCAYAL--LASGFDVWVESGL--KPYDFLSLVPVIEGAGGSITDWR 300
DB 186 KVR-----LRMCGSCALNLCGVACGRDLDFEYELFCGPDVAGGAVIVKEAGGFVDPDS 240
```

```
QY 301 GDKLHPVPTAESRPTSFNVVAAGDARVHKALDAL 335
DB 241 GSE--FOLTARR-----VAATNAHLKDFAIKAL 266
RESULT 7
SUHB_SALTY
ID SUHB_SALTY STANDARD; PRT; 267 AA.
AC P58537;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Inositol-1-monophosphatase (EC 3.1.3.25) (IMPase) (Inositol-1-
DE phosphatase) (I-1-Pase).
GN SUHB OR STM2546.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
RC MEDLINE=215334948; PubMed=11677609;
RA McLelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2.";
RL Nature 413:852-856(2001).
CC -!- CATALYTIC ACTIVITY: Myo-inositol 1-monophosphate + H(2)O -> myo-
CC inositol + phosphate.
CC -!- COFACTOR: Magnesium (By similarity).
CC -!- SIMILARITY: BELONGS TO THE INOSITOL MONOPHOSPHATASE FAMILY.
CC
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AF008815; AAL21440.1; -.
CC StyGene; SG72722; subh.
CC InterPro: IPR000760; Inositol_P.
CC PROSITE: PS00629; IMP_1; 1.
CC PROSITE: PS00630; IMP_2; 1.
CC HydroLase; Magnesium; Complete proteome.
CC KW SEQUENCE 267 AA; 29158 MW; FC46D8A8298BB334 CRC64;
Query Match 15.1%; Score 262; DB 1; Length 267;
Best Local Similarity 30.0%; Pred. No. 3.5e-13;
Matches 80; Conservative 44; Mismatches 123; Indels 20; Gaps 7;
QY 78 LVEAQAADAAGEVLKRYFR--ORVEIIDKEDHSPVTIADREAEAMVSVILKSFFT 135
DB 4 MLTIIVRAARKAGNVIKAKNYETPDVAESQKSGNSDFVTNVDKAAEAVIIDTIRKSPQHT 63
QY 136 IFGEENGWRC--ENSADFVWVLDPIDGTSFKTFGLFGLIALLHNGKPVIGVIDQ 195
DB 64 IITEEGEHVGTQ--DVQWVIDPLDGTTFNFKRLPHFAVSIKRTGRTEVAVVYDPMRN 122
QY 196 ERWIGVDKOTTLNGOEISVRSNLLAQAYLYTTSPLHFEA--DAEDAFIVRNKVPVPL- 253
DB 123 ELFTATRGGAQLNGVRLRGSTARDLDGTLATGFP--FKAKQYATYINIIKLTTECA 180
QY 254 -----YGCDCYAYALLASGFPVDIVVESGLKPYDFLSLVPVIEGAGGSITDWRGDLHPVT 309
DB 181 DFRRTGSAALDLAYVAAGRVDGFEIGLRPWFDAAGELLVREAGGIVDSFTGCH--NYMMT 239
QY 310 AESRPTSFNVVAAGDARVHKALDALR 336
```

Db 240 GN-----IVAGNPRVVKAMLANMR 258

RESULT 8

SUBH_ECOLI STANDARD; PRT: 267 AA.

AC P22783; P77511; Q8X2B6;

DT 01-AUG-1991 (Rel. 19, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Inositol-1-monophosphatase (EC 3.1.3.25) (IMPase) (Inositol-1-phosphatase) (i-1-pase).

GN SUBH OR SSVA OR B2533 OR Z3800 OR ECS3399.

OS Escherichia coli, and

OS Escherichia coli O157:H7.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

OC Escherichia.

OX NCBI_TaxID=562, 83334;

[1]

SEQUENCE FROM N.A.

STRAIN=K12;

MEDLINE=90202735; PubMed=2138605;

Yano R., Negai H., Shiba K., Yura T.;

"A mutation that enhances synthesis of sigma 32 and suppresses temperature-sensitive growth of the rpoH15 mutant of Escherichia coli.";

J. Bacteriol. 172:2124-2130(1990).

[2]

SEQUENCE FROM N.A.

STRAIN=K12 / MG1655;

MEDLINE=97426617; PubMed=9278503;

Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V., Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F., Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J., Mau B., Shao Y.;

"The complete genome sequence of Escherichia coli K-12.";

Science 277:1453-1474(1997).

[3]

SEQUENCE FROM N.A.

STRAIN=K12;

MEDLINE=97349980; PubMed=9205837;

Yamanoto Y., Aiba H., Baba T., Hayashi K., Inada T., Isono K., Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsuhashi N., Mizobuchi K., Mori H., Nakade S., Nakamura Y., Nashimoto H., Oshino T., Oyama S., Saito N., Sampei G., Satoh Y., Sivasubram S., Tagami H., Takahashi H., Takeda J., Takemoto K., Uehara K., Wada C., Yamagata S., Horiuchi T.;

"Construction of a contiguous 874-kb sequence of the Escherichia coli K-12 genome corresponding to 50.0-68.8 min on the linkage map and analysis of its sequence features.";

DNA Res. 4:91-113(1997).

[4]

SEQUENCE FROM N.A.

STRAIN=O157:H7 / EDL933 / ATCC 700927;

MEDLINE=21074935; PubMed=11206551;

Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D., Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A., Posfai D.G., Hackett J., Klink S., Boutin A., Shao Y., Miller L., Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamocis K., Apodaca J., Anantharam T.S., Lin J., Yen G., Schwartz D.C., Welch R.A., Blattner F.R.;

"Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";

Nature 409:529-533(2001).

[5]

SEQUENCE FROM N.A.

STRAIN=O157:H7 / RIMD 0509952;

MEDLINE=21156231; PubMed=11258796;

Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K., Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T., Lida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T., Kuhara S., Shiba T., Hattori M., Shinagawa H.;

"Complete genome sequence of enterohemorrhagic E. coli O157:H7 and genomic comparison with a laboratory strain K-12.";

DNA Res. 8:111-22(2001).

[6]

FUNCTION.

MEDLINE=95095968; PubMed=8002619;

Matsuhisa A., Suzuki N., Noda T., Shiba K.;

"Inositol monophosphatase activity from the Escherichia coli subB gene product.";

J. Bacteriol. 177:200-205(1995).

[7]

CHARACTERIZATION.

MEDLINE=20213266; PubMed=10747806;

Chen L., Roberts M.F.;

"Overexpression, purification, and analysis of complementation behavior of E. coli subB protein: comparison with bacterial and archaeal inositol monophosphatases.";

RL Biochemistry 39:4145-4153(2000).

CC -|- CATALYTIC ACTIVITY: Myo-inositol 1-monophosphate + H(2)O -> myo-inositol + phosphate.

CC -|- COFACTOR: Magnesium.

CC -|- SUBUNIT: Monomer.

CC -|- SIMILARITY: BELONGS TO THE INOSITOL MONOPHOSPHATASE FAMILY.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC

CC ENBL; M34828; AAA67506.1; -.

DR ENBL; AF000339; AAC75586.1; -.

DR ENBL; D90883; BAA16427.1; -.

DR ENBL; D90884; BAA16435.1; -.

DR ENBL; AE003483; -; NOT_ANNOTATED_CDS.

DR ENBL; AF002562; BAB36822.1; -.

DR PIR; A35158; A35158.

DR HSSP; P29218; 1IMF.

DR SWISS-2DPAGE; P22783; COLI.

DR EcoGene; EGI0983; subB.

DR InterPro; IPR000760; Inositol_P.

DR Pfam; PF00459; inositol_P; 1.

DR PROSITE; PS00629; IMP_1; 1.

DR PROSITE; PS00630; IMP_2; 1.

KW Hydrolase; Magnesium; Complete proteome.

FT CONFLICT 141 141

SQ SEQUENCE 267 AA; 29172 MW; 8FEC3508BD11301 CRC64;

Query Match 14.9%; Score 258; DB 1; Length 267;

Best Local Similarity 29.6%; Pred. No. 7.2e-13;

Matches 79; Conservative 44; Mismatches 124; Indels 20; Gaps 7;

Qy 78 LVEVAQRAADAAGEVLKRYFR--QVRVEIIDKHDSPTVIADREAEAMVSVILKSFPTHA 135

Db 4 MLNIAVRAARKAGNLIANKYETPDVAEASQKGSNDFTVNDKAAEAVIIDTIRKSPQHT 63

Qy 136 IFGEENGWRCNAENSADFVWVLDPIDGTFKSFITCKPLFGTLIALHNGKPVIGVIDQPILR 195

Db 64 IITEESG-ELEGTDQDVQWVIDPLDGTTFIKRLPHFAVSIARVIGKRTVEAVVYDPMRN 122

Qy 196 ERWIGVDGKQTTLLQGEISVRSNCLLAQAYLVYTSPLHFEA-DAEDAFIRVNKV----- 249

Db 123 ELTATRGQGAQLNGYRLRCSTARDLDGTLATGFP--FKAQYATTYINIVGKLFNECA 180

Qy 250 KVPYGCDCYAYALLASGFVDIVVESGLKPYDFLSLVPIEGAGGSITDWRGDKLHPVPT 309

Db 181 DFRRTGSAALDLAYVAAGRVDGFEIGLRPWFDAACELLVREAGGIVSFTGGH-NYMLT 239

Qy 310 AESRPTSFNVVAAGDARVHKEDALDLR 336

Db 240 GN-----IVAGNPRVVKAMLANMR 258

```
RESULT 9
SUHB_VIBCH STANDARD; PRT; 267 AA.
AC O9KTY5;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Inositol-1-monophosphatase (EC 3.1.3.25) (IMPase) (Inositol-1-
DE phosphatase) (i-1-Pase).
GN VC0745.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
RN [1]
SEQUENCE FROM N.A.
RP STRAIN=El Tor N16961 / Serotype O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Emolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae."
RL Nature 406:477-483(2000).
CC -1- CATALYTIC ACTIVITY: Myo-inositol 1-monophosphate + H(2)O = myo-
CC inositol + phosphate.
CC -1- COFACTOR: Magnesium (By similarity).
CC -1- SIMILARITY: BELONGS TO THE INOSITOL MONOPHOSPHATASE FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; AE004160; AAF93910.1; ALT_INIT.
CC HSSP; P29218; 11MF.
CC TIGR; VC0745;
DR InterPro: IPR000760; Inositol_P.
DR Pfam; PF00459; inositol_P; 1.
DR PROSITE; PS00629; IMP_1; 1.
DR PROSITE; PS00630; IMP_2; 1.
DR Hydrolyase; Magnesium; Complete proteome.
KW SEQUENCE 267 AA; 29088 MW; 88AD8CDF78A7AD87 CRC64;
-----
Query Match 14.3%; Score 248; DB 1; Length 267;
Best Local Similarity 28.9%; Pred. No. 4.2e-12;
Matches 79; Conservative 41; Mismatches 121; Indels 32; Gaps 8;
QY 78 LVEVAQRAADAAGEVLKRYF--QRVEIDKEDHSPVTIADREAEAMSVILKSPPTA 135
DB 4 MLNIAIARAARKAGNHIAKSLAENAEKIQTTQKGSNDFVTVNVDKAEALIVSTIKSYPEHC 63
QY 136 IFGEENGWRCNAENSADFVWVLDPIDGTSFKITGKPLFGTLIALHNGKPVIGIDOPILR 195
DB 64 IIAEEG-LLEGDKKEQVNIIDPLDGTNTNPKGFPFANVSIAVFRFGKTEVACVYDPMN 122
QY 196 ERWIGVDGKOTTLNGOEISVRSNLLAQAYLYTTSPLHFEADAEAFIRVRNKVKPLYG 255
DB 123 ELETAQRGAGQAQLNARIRVQIKDLOGAVLATAPFPKQKOHSE-SFMKILSAMFEV--- 178
QY 256 CD-----CYAYALLASGFVDIVVESGLKPYDFLSLVPVIEGAGGSIIDWRGDK 303
DB 179 CADFRRTGSAALDLCY-----LAANRVGDYFELGLKPMDMAAGELIAREAGAIVTFDAG-- 232
QY 304 LHPVPTAESRPTSFNVAAAGDARVHKALDALR 336
-----
Db 233 -----GTDYMQSGNIVASSPRGV-KAILQHIR 258
RESULT 10
MYOL_LYCES STANDARD; PRT; 273 AA.
ID MYOL_LYCES AC P54926;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myo-inositol-1(or 4)-monophosphatase 1 (EC 3.1.3.25) (IMPase 1) (IMP
DE 1) (Inositol monophosphatase 1).
GN IMP1.
OS Lycopersicon esculentum (Tomato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
OX NCBI_TaxID=4081;
RN [1]
SEQUENCE FROM N.A.
RP STRAIN=cv. VENT Cherty;
RX MEDLINE=96351935; PubMed=8718627;
RA Gillaspay G.E., Keddle J.S., Oda K., Gruissem W.;
RA "Plant inositol monophosphatase is a lithium-sensitive enzyme encoded
RA by a multigene family."
RT Plant Cell 7:2175-2185(1995).
RL
CC -1- FUNCTION: IT IS RESPONSIBLE FOR THE PROVISION OF INOSITOL REQUIRED
CC FOR SYNTHESIS OF PHOSPHATIDYLINOSITOL AND POLYPHOSPHOINOSITIDES.
CC -1- CATALYTIC ACTIVITY: Myo-inositol 1-monophosphate + H(2)O = myo-
CC inositol + phosphate.
CC -1- COFACTOR: REQUIRES MAGNESIUM (BY SIMILARITY).
CC -1- ENZYME REGULATION: INHIBITED BY LI(+).
CC -1- PATHWAY: KEY ENZYME OF THE PHOSPHATIDYL INOSITOL SIGNALING
CC PATHWAY.
CC -1- SIMILARITY: BELONGS TO THE INOSITOL MONOPHOSPHATASE FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; U39444; AAB19030.1;
CC HSSP; P29218; 11MF.
CC InterPro: IPR000760; Inositol_P.
CC Pfam; PF00459; inositol_P; 1.
CC PROSITE; PS00629; IMP_1; 1.
CC PROSITE; PS00630; IMP_2; 1.
CC Hydrolyase; Magnesium; Lithium; Multigene family.
KW SEQUENCE 273 AA; 29558 MW; 92DID532E359E0FD CRC64;
-----
Query Match 14.3%; Score 247.5; DB 1; Length 273;
Best Local Similarity 28.9%; Pred. No. 4.8e-12;
Matches 80; Conservative 53; Mismatches 111; Indels 33; Gaps 10;
QY 76 ERLEVEVAQRAADAAGEVLKRYFQRVVEIDKEDHSPVTIADREAEAMSVILKSPPTA 135
DB 8 EELGVAVDAAKRAGEILIRKGFHETKRVHVKQVDLVETDKACEDLIFNHLKQHFPSHK 67
QY 136 IFGEENGWRCNAENSADF-----VWVLDPIDGTSFKITGKPLFGTLIALHNGKPVIGI 189
DB 68 FIGEET-----SAATGDFDLTDEPTWIDVDPDGTNTNFKVGFPSVCVSGITGKPTVGV 123
QY 190 DQILERWIGVDGKOTTLNGOEISVRSNLLAQAYL-----YTTSPHLFEADAEAFIRV 245
DB 124 YDPIDELFTFGINGKAYLNGRKPIKVSQSELSKSLGLTEVGTTRDNL---TVETTFRRI 180
QY 246 RN---KVKVPLVGCDCYAYAL---LASGFVDIVVESGL-KPYDFLSLVPVIEGAGGSIIDW 299
DB 181 NNLLIKVR-SLRMCGSCALDLCWVACGRLEFLYLGYPWDVAGGAVIVKEAGGVLFDP 239
```



```
Db 249 TRIAASNP 256      | | | |
RESULT 13
SUBH_XYLFA                STANDARD:      PRT: 275 AA.
AC Q9PAM0;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Inositol-1-monophosphatase (EC 3.1.3.25) (IMPase) (Inositol-1-
DE phosphatase) (I-1-Pase).
GN SUHB OR XF2476.
OS Xylella fastidiosa.
OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
OC Xylella.
OX NCBI_TaxID=2371;
RN [1]
SEQUENCE FROM N.A.
C STRAIN=9a5C;
X MEDLINE=20365717; PubMed=10910347;
RA Simpson A.J.G., Reinach F.C., Artuda P., Aureu F.A., Acencio M.,
RA Alvaranga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrier H.,
RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,
RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA Peixoto R.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
RA Quadagno R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsubako M.H.,
RA Vallada H., Van Sluys M.A., Verjowski-Almeida S., Vettore A.L.,
RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
RT "The genome sequence of the plant pathogen Xylella fastidiosa.";
RL Nature 406:151-159(2000).
CC -!- CATALYTIC ACTIVITY: Myo-inositol 1-monophosphate + H(2)O = myo-
CC inositol + phosphate.
CC -!- COFACTOR: Magnesium (By similarity).
CC -!- SIMILARITY: BELONGS TO THE INOSITOL MONOPHOSPHATASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE004055; AAF85274.1; -.
DR HSSP: P29218; 2HHM.
DR InterPro: IPR000760; Inositol_P.
DR Pfam: PF00459; inositol_P; 1.
DR PROSITE: PS00629; IMP-1; 1.
DR PROSITE: PS00630; IMP-2; 1.
KW Hydrolase; Magnesium; Complete proteome.
SQ SEQUENCE 275 AA: 30233 MW; 74F622ACC4EE082C CRC64;

Best local Similarity 28.4%; Pred. No. 5.2e-11;
Matches 73; Conservative 51; Mismatches 117; Indels 16; Gaps 8;

Qy 79 VEVAQRADAAGEVLRKYFR--QRVEIIDKEDHSPVTIADREAEAEAMVSVTLKSPFTHAI 136
Db 6 VNIMVKAARSAGNVLRLHINKLETLHWIQKSRMDYASDVDEMAEKVIVKEKRAYPEYGI 65
Qy 137 FGEENGWRCAENSADFVWVLDPIDGTSFTGKPLFTGLIALLHNGKPVIGVIGDIPILRE 196
Db 66 LGEEGG---LQGNHRIMWIDPLDGTSNYLRGPHYCISIALVENGEPTDAVFDPLRNE 122
Qy 197 RWIGVDGKQTLNCOFIISVRSCNIIAQAYLYT-TSP-HLFEDAE---DAFIRVRNKVK 250
Db 123 LFTASRGAGAILNPKIRKVANRKDLNGLMTNLTGFSRPSRAHAQLKCVDAALLMQAEDIR 182
Qy 251 VPLYGCDYAYALLASGFVDIIVESGLKPYDFLSLVPVIEGAGGSITDWRGDKLRHPVTA 310
Db 183 --RSGSRAALDLAYVACGRADAYFEAGIKVWDVAAGMLLVREAGGYVCDFFKGADA--PRMD 238
Qy 311 ESRPTSFNVVAAGDARV 327
Db 239 DKGPESQVLV-AGNIKV 254

RESULT 14
SUHB_THEMEA                STANDARD:      PRT: 256 AA.
AC O3832;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Inositol-1-monophosphatase (EC 3.1.3.25) (IMPase) (Inositol-1-
DE phosphatase) (I-1-Pase).
GN SUHB OR TM1415.
OS Thermotoga maritima.
OC Bacteria; Thermotogae; Thermotogae (class); Thermotogales;
OC Thermotogaceae; Thermotoga.
OX NCBI_TaxID=2336;
RN [1]
SEQUENCE FROM N.A.
RP STRAIN=MSB8 / DSM 3109;
RX MEDLINE=98386670; PubMed=9720201;
RA Liebl W., Brem D., Gotschlich A.;
RT "Analysis of the gene for beta-fructosidase (invertase, inulinase) of
RT the hyperthermophilic bacterium Thermotoga maritima, and
RT characterisation of the enzyme expressed in Escherichia coli.";
RL Appl. Microbiol. Biotechnol. 50:55-64(1998).
RN [2]
SEQUENCE FROM N.A.
RP STRAIN=MSB8 / DSM 3109;
RX MEDLINE=99287316; PubMed=10360571;
RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT "Evidence for lateral gene transfer between Archaea and Bacteria from
RT genome sequence of Thermotoga maritima.";
RN Nature 398:323-329(1999).
RN [3]
CHARACTERIZATION.
RX MEDLINE=99437848; PubMed=10508089;
RA Chen L., Roberts M.F.;
RT "Characterization of a tetrameric inositol monophosphatase from the
RT hyperthermophilic bacterium Thermotoga maritima.";
RL Appl. Environ. Microbiol. 65:4559-4567(1999).
CC -!- FUNCTION: Displays a 20-fold higher rate of hydrolysis of the D
CC isoform of inositol 1-phosphate than of the L isoform.
CC -!- CATALYTIC ACTIVITY: Myo-inositol 1-monophosphate + H(2)O = myo-
CC inositol + phosphate.
CC -!- COFACTOR: Magnesium.
CC -!- SUBUNIT: Homotetramer.
```


R; Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, H.
 erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClellan
 : Karp, P.; Romero, P.; Zhang, S.
 Science 294, 2317-2323, 2001
 A: Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
 ster, E.W.
 A: Title: The Genome of the Natural Genetic Engineer *Agrobacterium tumefaciens* C58.
 A: Reference number: AB2577; PMID: 11743193
 A: Accession: AG3034
 A: Status: preliminary
 A: Molecule type: DNA
 A: Residues: 1-262 <KUR>
 A: Cross-references: GB:AE008689; PIDN: AAL44693.1; PID: gl7742322; GSPDB: GN00187
 A: Experimental source: strain C58 (Dupont)
 C: Genetics:
 A: Gene: Atu3895
 A: Map position: linear chromosome
 C: Superfamily: suppressor protein subh

T..3.

Query Match 30.3%; Score 525; DB 2; Length 262;
 Best Local Similarity 42.7%; Pred. No. 7.3e-34;
 Matches 112; Conservative 51; Mismatches 73; Indels 26; Gaps 7;

Qy 84 RAADAAGEVLKRYRQRIIDKED--HSPVTIADREAEAMVSVILKSPFTHAIFGEEN 141

Db 14 KLADAASAEPLRPTGTAVINKQGGYDPVTEGDQAAETAIRALIEERFPQHILGEEH 73

Qy 142 GWRCAEN---SADPVMVLDPIDGTSFKITGKPLFGTLIALLHNGKPVIGVIDOPILRERW 198

Db 74 G-----NVGLDRDHIWIDPIDGTRAFISGVPMWGTGLIGFOSSGRATMGINDOPFTTKRY 128

Qy 199 IGVDGKQTTLNG---OESIVRSNLLAQAYLYTTSPHLFEADAEDAFIRVRNKVKVPLY 254

Db 129 F-ADGKAANYFGPDGKKIRTRDCASLSDAVLTFTTTHIETAEKPLYEKVQDQVRLFRY 187

Qy 255 GDCYAYALLASGFVDIVVESGLKPYDFLSLVPIEGAGSGITDWRGDKLHWPVTAESRP 314

Db 188 GVCYAYCLLAAGHVDLVIESGLKPYDVGALIPVIEQAGGTMTWDG-----GRP 237

Qy 315 -TSFNVAAGDARVHKEALDAL 335

Db 238 ENGRILAAAGSKAVHEALAIL 259

RESULT 3

Similar to mlcbl779.29 (probable monophosphatase gene) in mycobacterium leprae [imported]

Species: *Agrobacterium tumefaciens*

Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 15-Feb-2002

Accession: E98251

R: Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quorollo, B.; Goldman,

A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;

Science 294, 2323-2328, 2001

A: Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent *Agrobacterium tum*

A: Reference number: A97359; PMID: 11743194

A: Accession: E98251

A: Status: preliminary

A: Molecule type: DNA

A: Residues: 1-297 <KUR>

A: Cross-references: GB:AE007870; PIDN: AAK89535.1; PID: gl5159414; GSPDB: GN00170

C: Genetics:

A: Map position: linear chromosome

C: Superfamily: suppressor protein subh

Query Match 30.3%; Score 525; DB 2; Length 297;

Best Local Similarity 42.7%; Pred. No. 8.4e-34;

Matches 112; Conservative 51; Mismatches 73; Indels 26; Gaps 7;

Qy 84 RAADAAGEVLKRYRQRIIDKED--HSPVTIADREAEAMVSVILKSPFTHAIFGEEN 141

Db 14 KLADAASAEPLRPTGTAVINKQGGYDPVTEGDQAAETAIRALIEERFPQHILGEEH 73

Db 49 KLADAASAEPLRPTGTAVINKQGGYDPVTEGDQAAETAIRALIEERFPQHILGEEH 108

Qy 142 GWRCAEN---SADPVMVLDPIDGTSFKITGKPLFGTLIALLHNGKPVIGVIDOPILRERW 198

Db 109 G-----NVGLDRDHIWIDPIDGTRAFISGVPMWGTGLIGFOSSGRATMGINDOPFTTKRY 163

Qy 199 IGVDGKQTTLNG---OESIVRSNLLAQAYLYTTSPHLFEADAEDAFIRVRNKVKVPLY 254

Db 164 F-ADGKAANYFGPDGKKIRTRDCASLSDAVLTFTTTHIETAEKPLYEKVQDQVRLFRY 222

Qy 255 GDCYAYALLASGFVDIVVESGLKPYDFLSLVPIEGAGSGITDWRGDKLHWPVTAESRP 314

Db 223 GVCYAYCLLAAGHVDLVIESGLKPYDVGALIPVIEQAGGTMTWDG-----GRP 272

Qy 315 -TSFNVAAGDARVHKEALDAL 335

Db 273 ENGRILAAAGSKAVHEALAIL 294

RESULT 4

F95293

probable inositol monophosphatase [imported] - *Sinorhizobium meliloti* (strain 1021) n

C: Species: *Sinorhizobium meliloti*

C: Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001

C: Accession: F95293

R: Barnett, M.J.; Fisher, R.F.; Jones, T.; Komp, C.; Abola, A.P.; Barloy-Hubler, F.; F

; Kalman, S.; Keating, D.H.; Palm, C.; Peck, M.C.; Surzycki, R.; Wells, D.H.; Yeh, K

Proc. Natl. Acad. Sci. U.S.A. 98, 9883-9888, 2001

A: Title: Nucleotide sequence and predicted functions of the entire *Sinorhizobium meli*

A: Reference number: A95262; MUID: 21396509; PMID: 11481432

A: Accession: F95293

A: Status: preliminary

A: Molecule type: DNA

A: Residues: 1-108 <KUR>

A: Cross-references: GB:AE006469; PIDN: AAK64912.1; PID: gl4523333; GSPDB: GN00165

A: Experimental source: strain 1021, megaplasmid pSymA

R: Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubl

pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.

L.; Hyman, R.W.; Jones, T.

Science 293, 668-672, 2001

A: Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lela

heault, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh,

A: Title: The composite genome of the legume symbiont *Sinorhizobium meliloti*.

A: Reference number: A96039; MUID: 21368234; PMID: 11474104

A: Contents: annotation

C: Genetics:

A: Gene: SMA0483

A: Genome: plasmid

Query Match 27.9%; Score 483; DB 2; Length 308;
 Best Local Similarity 39.6%; Pred. No. 1.8e-30;
 Matches 107; Conservative 38; Mismatches 103; Indels 22; Gaps 4;

Qy 70 VGEATERLVEVQRAADAGEVLKRYRQRIIDKEDHSPVTIADREAEAMVSVILK 129

Db 56 LGEPAS-----FAHDIADIAHQATISSAAGVRPEIAKSDASPVETDRAVEKCLRRIAD 110

Qy 130 SFPTHAIFGEENGWRCFAENSADFVWVLDPIDGTSFKITGKPLFGTLIALLHNGKPVIGVI 189

Db 111 HFPDHGVLGEEFGAEGUGN--EFVWVDPIDGTRAFVAGLVPVYGTLSLFRGGTPIUGLI 168

Qy 190 DQILRERWIGVDGKQTTLNGOEISVRSNLLAQAYLYTTSPHLFEADAEDAFIRVRNKV 249

Db 169 DNPMTGDRMLGVSGOPTLLNNVPRTASTATATAFANGNPDAFSPADKSRVSLRRIT 228

Qy 250 KVPYLGDCYAYALLASGFVDIVVESGLKPYDFLSLVPIEGAGSGITDWRGDKLHWPVT 309

Db 229 RWCYVYGCYAYGRVADGSDISIDGGLDPYDYCALVPVITGAGCITDQOG----- 280

Qy 310 AESRPTSFN-----VVAAGDARVHKEALDAL 335

Db 281 ---RPLTNSGGLCVATATDCLLHRRVLEIL 307

RESULT 5
C87528
inositol monophosphatase family protein [imported] - Caulobacter crescentus
C:Species: Caulobacter crescentus
C>Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C:Accession: C87528
R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A:Title: Complete Genome Sequence of Caulobacter crescentus.
A:Reference number: A87249; MUID:21173698; PMID:11259647
A:Accession: C87528
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-278 <STO>
A:Cross-references: GB:AE005673; NID:g13423761; PIDN:AAK24223.1; GSPDB:GN00148
Genetics:
.:Gene: CC2252

Query Match 27.6%; Score 477; DB 2; Length 278;
Best Local Similarity 40.8%; Pred. No. 4.7e-30;
Matches 116; Conservative 48; Mismatches 88; Indels 32; Gaps 9;
QY 73 LATERLVEAQAAD---AAGEVLKRYRQRYVEIIDK-----EDHS---PVTIADRE 118
DB 3 LSADRLAALDAFILDNLNRSASVILFLFRADHGLEDKAGKNLPDRDTHAAFPDPTVTEADRG 62
QY 119 ABEAMSVILKSPFTHAIFGEENGWRCARN--SADFVWVLDPIDGTSFKITKPLFGTLI 176
DB 63 AEAATRALIAQRPDPHGVIIEYG---EDRPDAEFVWLDPIDGTRAFISGLPWGTLVLT 118
QY 177 ALLHNGKPVIGVIDPILRERWIGVGKQTTLNGQE---ISVRSCNLLAQAYLYTTSPLH 232
DB 119 GLRHEGRPLVLSIGTPYVNEIFIGHAGGARLVSGGEARPIRVRECANINDAVIATTDPA 178
QY 233 LREADAEDAFIRVNKKVPLVGCDCYAYALLASGFVDIVVESGLKPYDFLSLVPVIEGA 292
DB 179 CRDGAERGAWLQVRAAKIARLGCDAIYAMVAMGKMDMVIIEAGLUKSWDIEAIPLEIGA 238
QY 293 GGSITDWRGDKLHWPVPTAESRPTSFNVVAAGDARVHKEALDALR 336
DB 239 GGVNTNRGCE---PV---GFGNGGVISGDRRPLDEALVSLR 274

RESULT 6
AF3515
extragenic suppressor protein subh [imported] - Brucella melitensis (strain 16M)
C:Species: Brucella melitensis
C>Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 15-Feb-2002
C:Accession: AF3515
R:DeIvecchio, V.G.; Kapatal, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova,
.; Mazur, M.; Goltzman, E.; Selkov, E.; Elizer, P.H.; Hagius, S.; O'Callaghan, D.; Letess
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A:Title: The genome sequence of the facultative intracellular pathogen Brucella melitensis
A:Reference number: AD3252; PMID:11756688
A:Accession: AF3515
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-272 <KUR>
A:Cross-references: GB:AE008918; PIDN:AAL53289.1; PID:g17984172; GSPDB:GN00191
A:Experimental source: strain 16M
C:Genetics:
A:Gene: BMEI10048
A:Map position: 11
C:Superfamily: suppressor protein subh

Query Match 26.8%; Score 463.5; DB 2; Length 272;
Best Local Similarity 41.3%; Pred. No. 5.3e-29;

Matches 116; Conservative 47; Mismatches 93; Indels 25; Gaps 9;
QY 66 RAAAVGELATERLVEVA---QRAADAAGEVLKRYRQRYVEIIDKED--HSPVTIADREAE 120
DB 4 RAPCGNRRRTGLLIDKAFSEVAAAAAQTLPF--FRLQTEVDNKNYSVGDFPVTADRAAE 62
QY 121 EAMSVILKSPFTHAIFGEENGWRCARN--SADFVWVLDPIDGTSFKITKPLFGTLIALL 179
DB 63 RAIRAVIGRTFFPHGILGEEYG---AENTDRSHVWIIDPDGTRAFISGLPWGTLVGLT 119
QY 180 HNGKPVIGVIDPILRERWIGVGKQTTLN----GQEISVRSCNLLAQAYLYTTSPLHF 234
DB 120 VGDARAGMMSQPTTCEIFYS--DGDGAYLQDGDGAPRRLCVKRNAVLEADATLFTTTTALF 178
QY 235 EADAEDAFIRVNKKVPLVGCDCYAYALLASGFVDIVVESGLKPYDFLSLVPVIEGAGG 294
DB 179 KGDRAKAFDRLESASVRLSRVYDCAFAFAMLAGGFVDIVVEAGLQTYDIAALIPILBQAGG 238
QY 295 SITDWRGDKLHWPVPTAESRPTSFNVVAAGDARVHKEALDAL 335
DB 239 VVTRRDGG-----PAHQ---CGDIVAAATPALHQAALL 270

RESULT 7
S74899
extragenic suppressor - Synecocystis sp. (strain PCC 6803)
N:Alternate names: protein s111959
C:Species: Synecocystis sp.
A:Variety: PCC 6803
C>Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 20-Jun-2000
C:Accession: S74899
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima,
O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Ya
DNA Res. 3, 109-136, 1996
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synecocystis
S.
A:Reference number: S74322; MUID:97061201
A:Accession: S74899
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-267 <KAN>
A:Cross-references: EMBL:D90909; GH:AB001339; NID:g1652844; PIDN:BAAL7860.1; PID:g16
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C:Superfamily: suppressor protein subh

Query Match 22.9%; Score 496.5; DB 2; Length 267;
Best Local Similarity 36.9%; Pred. No. 9.6e-24;
Matches 101; Conservative 49; Mismatches 59; Indels 25; Gaps 10;

QY 76 ERLVEVAORAADAAGEVLKRYRQRYVEIIDKEDHSPVTIADREAEAMSVILKSP 131
DB 6 EORLFIAQQLAAVSGTEILFYFRSHLQGGTKIDQVS--ALVTQADEAEQAMVDLIQAQF 64
QY 132 PTHAIFGEENGWRCARN--SADFVWVLDPIDGTSFKITKPLFGTLIALLH--NGKPVIGVID 190
DB 65 PQDGVITREE--GKNIAKSG--YTWVLDPIDGTSFVRGLPIEATLIGLVDADMRLVGLIAH 122
QY 191 QPILRERWIGVGKQTTLNG-----QEISVRSCNLLAQAYLYTTSPLHFEADA--EDAFIRV 245
DB 123 QPISGRWQGVQGVQGSNVNGIPLVNPYKASEINLTA--ACIVSTPLMFTTPVQQQKMA 181
QY 246 RNKKVPLVGCDCYAYALLASGFVD---IVVESGLKPYDFLSLVPVIEGAGGSITDWRGD 302
DB 182 YRQCQRTAFGGDCFNFLSASGWTAMPLVIVEADLNFYDFCALIPIITGANCYCTDQGG 241
QY 303 KLHWPVPTAESRPTSFNVVAAGDARVHKEALDALR 336
DB 242 EL-----TPESTEVVASPNPKLHSEILAFLO 267

RESULT 8
AC0353

probable inositol monophosphatase family protein YP02899 [imported] - Versinia pestis (s
C:Species: Versinia pestis
C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 09-Nov-2001
C:Accession: AC0353
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
Id, M.; Rutherford, K.; Simmonds, T.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrall,
Nature 413, 523-527, 2001
A:Title: Genome sequence of Versinia pestis, the causative agent of plague.
A:Reference number: AB0001; MUID:21470413; PMID:11586360
A:Accession: AC0353
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-267 <KUR>
A:Cross-references: GB:AL590842; PIDN:CAC92150.1; PID:g15980865; GSPDB:GN00175
C:Genetics:
A:Gene: YP02899
C:Superfamily: suppressor protein subh

Query Match 16.9%; Score 293; DB 2; Length 267;
Best Local Similarity 31.6%; Pred. No. 1.3e-15;
Matches 86; Conservative 43; Mismatches 111; Indels 32; Gaps 8;
Qy 78 LVEAQAADAAGEVLRKYFR--QRVEIIDKEDHSPVTIADREAEAMVSVILKSFPTA 135
Db 4 MLTIAIRAARKAGNLIANKYETPDVAESAKSGNDFTNVDRDAEHLIIDVIRKSPKHT 63
Qy 136 IFGEENGWRCANSAFVWVLDIDTGTGKSFITGKPLFTLTIALHNGKPVIGVIDQPILR 195
Db 64 IISECGELVGEDD-DVQWVIDPLDGTGTFNFKLPFHFVAVSIAVRKIGTFAVAVYDPMRN 122
Qy 196 ERWIGVGDCKOTTLNGOEISVRSCN-----LLAQAYLYTTSPLFEADAEDAFIRVKN 248
Db 123 ELFTASRGCGAQLNGYRL--RGTNAKDLDTGLLATGFFKVKQH-----APAYIRVVVK 174
Qy 249 V-----KVPLXGCCYAYALLASGFVDIVVESGLKPYDFLSLVVPVIEGAGGSITDWRGDK 303
Db 175 LFEQCADFRGTSGAALDLAYVAAGRVDFEFGELGKPMDFAGGELLVRESGGIVTDFAGGH 234
Qy 304 LHWVPTAESRPTSFNVVAAGDARVHKEALDAL 335
Mb 235 NHF-----SSGNIV-AGNPRIKVSIVQAM 257

RESULT 9
C83169
extragenic suppressor protein Subh PA3818 [imported] - Pseudomonas aeruginosa (strain PA
Species: Pseudomonas aeruginosa
Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
Accession: C83169
R:Stover, C.K.; Pham, X.O.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; B
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
; Iori, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic patho
A:Reference number: AB2950; MUID:20437337
A:Accession: C83169
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-271 <STO>
A:Cross-references: GB:AE004799; GB:AE004091; NID:g9949981; PIDN:AAG07205.1; GSPDB:GN001
A:Experimental source: strain PAO1
C:Genetics:
A:Gene: PA3818
C:Superfamily: suppressor protein subh

Query Match 16.9%; Score 292; DB 2; Length 271;
Best Local Similarity 29.7%; Pred. No. 1.6e-15;
Matches 82; Conservative 54; Mismatches 106; Indels 34; Gaps 8;
Qy 78 LVEAQAADAAGEVLRKYFR--QRVEI---DKEDHSPVTIADREAEAMVSVILKSPH 142

Db 4 MLNTALRAARSAGELJ---FRSIERLDVISYNEKDAKDYVTEVDRAAEQITVAALRKAYP 60
Qy 133 THAIFGEENGW-RCANSAFVWVLDIDTGTGKSFITGKPLFTLTIALHNGKPVIGVIDQ 191
Db 61 THAIFGEENGW-RCANSAFVWVLDIDTGTGKSFITGKPLFTLTIALHNGKPVIGVIDQ 191
Qy 192 PILRERWIGVGDCKOTTLNGOEISVRSCNLLAQAYLYTTSPLH-FEADAEDAFIR-----V 245
Db 121 PVROEFTASRGRGAALNRRRLRVSGRKSLEGALLGTGFFPRDNOIDNLDNYLNNFRSLV 180
Qy 246 RNKVKVPLVYGCDCYAYALIASGFVDIVVESGLKPYDFLSLVVPVIEGAGGSITDWRG---- 301
Db 181 GOTAGIRHAGAASIDILAYVAAGRYDATWFGISEWMAAGALLVQEAAGGLVSDFTGSHEF 240
Qy 302 -DKLHWVPTAESRPTSFNVVAAGDARVHKEALDALR 336
Db 241 LKRGH-----IVAGNTKCFKALJTTT 262

RESULT 10
T45317
monophosphatase [imported] - Mycobacterium leprae
C:Species: Mycobacterium leprae
C:Date: 31-Jan-2000 #sequence_revision 31-Jan-2000 #text_change 18-Feb-2000
C:Accession: T45317
R:Parkhill, J.; Barrall, B.G.; Rajandream, M.A.
Submitted to the EMBL Data Library, August 1997
A:Reference number: 222864
A:Accession: T45317
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-255 <PAR>
A:Cross-references: EMBL:Z98271; PIDN:CAB11010.1
A:Experimental source: cosmid H1779
C:Genetics:
A:Note: MLCB1779.29
C:Superfamily: suppressor protein subh

Query Match 16.4%; Score 283; DB 2; Length 255;
Best Local Similarity 31.4%; Pred. No. 7.6e-15;
Matches 85; Conservative 40; Mismatches 102; Indels 44; Gaps 9;
Qy 81 VAQRAADAAGEVLRKYFR--KQVETIDKEDHSPVTIADREAEAMVSVILKSFPTHAIRGE 139
Db 2 LALTLAIDRAKADALISARRKALNIRVY--TVTHADRAVFAVDVRAVLGRERPKDGILGE 61
Qy 140 FNGWRCANSAFVWVLDIDTGTGKSFITGKPLFTLTIALHNGKPVIGVIDQPILRERW- 198
Db 62 EYGGTITPSSGQQ--WIVVPIUGTKNFYKGVPMWASLIALLLEDGVPSIGVWSAPALQRW 119
Qy 199 -----IGVVG-----KOTTLNGOEISVRSCNLLAQAYLYTTSPLHFEADAEDAF 242
Db 120 AARGOCAFVAVDVGPVKRIAVSEVADINSASISFSLSGWAQRL-----RDRF 167
Qy 243 IRVRNKV-KVPLVYGCDCYAYALIASGFVDIVVESGLKPYDFLSLVVPVIEGAGGSITDWRG 301
Db 168 LELTDVAVRVKAYG--DFISYGLLAEGAIDVAAPKVSVDLAALDIVVREAGGVLTGLDG 226
Qy 302 DKLHWVPTAESRPTSFNVVAAGDARVHKEAL 332
Db 227 -----TPPHGGSAVAT-NGRLHQEVL 247

RESULT 11
C70470
myo-inositol-1(or 4)-monophosphatase - Aquifex aeolicus
C:Species: Aquifex aeolicus
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
C:Accession: C70470
R:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.;
V.

	Query Match	15.8%	Score 273.5	DB 2	Length 260
	Best Local Similarity	32.5%	Pred. No. 4.4e-14		
	Matches 88	Conservative 42	Mismatches 104	Indels 37	Gaps 11
Qy	81	VAQRAADAAGVLNR-KYFQRQREIIDKHDSPTVIADREAEEAMVSVILKSFPTHAFGE	139		
		: : : : : : : : :			
Db	8	LALALADRADELTRVFCALDLRITDPDLTPVTDADRANVESDVRLGRDRPDCGVLCGE	67		
Qy	140	ENGWRCAENSADFV---WVLDPIDGTSFSTGKPLFCGI	196		

Qy	78	LVEVAQRAADAAGEVLKRYQR--VEIDKEHDSPVYTIADREAEEMWVILKSPPTHA	135
Db	4	MLNIIAARAARKAGNVIKAKNYERRDAIESTQKINDVYTNVDKASEAEIIIEIRKSPYDHT	63
Qy	136	IFGSENCWRCAENSADPFWVLDPIHDGTKSFITCKPLFGTLTIALHNGCKPVIGVIOPILR	195
Db	64	LIITEETG-AIEGKUSDUVQWIIDPLDCTHFNWGTGLPHEVSIVAVRKNRTEVGVVYDPIRN	122
Qy	196	ERWIGVDCQKOTTLNGQELIVSKSNLLAAQAYLYTTSPIH-----LFEADAEDAF-I	243
Db	123	ELFTAVRGEKAKLNEVRI.RVDSKREIQGSLATGTFPPKQKLPAPQAFMMNALIIEAADF	182
Qy	244	RVRNKVKVPLVGCDCYAYALLASGFVDIVVESGILKPYDFELSLVPIVBGSGSTIDNRGDK	303
Db	183	RTCSAALDL-----CY-----VASNRIDGYFEMGLKAWDCAAGDLIVREAGLVCDFD---	231
Qy	304	LHWPVTAEISRPTSNVVAAGDARVHKALDALR	336
Db	232	-----AGNSYLRSNIIIAA-PSRVIIKEMLNKIR	258

```

RESULT 14
T07795
myo-inositol-1(or 4)-monophosphatase (EC 3.1.3.25) isoform 3 - tomato
N:Alternate names: myo-inositol-1-phosphatase
C:Species: Lycopersicon esculentum (tomato)
C:Date: 14-May-1999 #sequence:revision 14-May-1999 #text_change 08-Oct-1999
R:Accession: T07795
R:Gillespy, G.E.; Keddle, J.S.; Oda, K.; Gruissem, W.
Plant Cell 7, 2175-2185, 1995
A:Title: Plant inositol monophosphatase is a lithium-sensitive enzyme encoded

```


[illegible]


```

; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/347,471
; FILING DATE: 08-DECEMBER-94
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5710015th, Robert J.
; REGISTRATION NUMBER: 27,366
; REFERENCE/DOCKET NUMBER: T-1160
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (908) 594-7262
; TELEX: (908) 594-4720
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 277 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-347-471-2

Query Match 11.2%; Score 193.5; DB 1; Length 277;
Best Local Similarity 25.5%; Pred. No. 8.9e-13;
Matches 70; Conservative 51; Mismatches 120; Indels 33; Gaps 10;

QY 76 ERLVEAQAADAAGEVLRKYFRQVEIIDKEDHSP---VTIADREAEAEAMVSVILKSP 132
DB 6 QECMDYAVTLARAGEVVCIAIKNEMNMLKS---SPVDLVATDQKVEKMLISSIKEKYP 63
QY 133 THAIFGEENGWRCANSA---DFWVWLDPIDGTGKSFITGKPLFGTLIAL' HNGKPVIGVI 189
DB 64 SHSFIGEES-VAAGEKSILTDNPTWIIDPIDGTTFNVRFPFVAVSIGI..NKKIEFGVV 122
QY 190 DQPIRLERWTGVDGKQTTLNGQEISVRSNLLAQAYLYTTSPLHFEADAEDAFIRVNKV 249
DB 123 YSCVEGKMVTARKGKGAFCNGKLOVSQQEDITKSLVT---ELGSSRTPTETVRMVLNNM 179
QY 250 K----VPLYGCCDCYAVA-----LLASCFYDIVVESGLKPYDFLSLVPVIEGAGSITDWR 300
DB 180 EKLFCEIPVHGIRSVGTAANNMCLVATGGADAYYEMGHCMWDVACAGIIVTEAGGVLM DVT 239
QY 301 GDLKHPVTAESRPTSFNVVAAGD---ARVHKE 330
DB 240 GG----PFDLMSR----RVIAANNRILAERIAKE 265

RESULT 9
US-09-091-952A-26
Sequence 26, Application US/09091952A
Patent No. 6458532
GENERAL INFORMATION:
APPLICANT: Detera-Wadleigh, Sevilla D.
Gershon, Elliot S.
Badner, Judith A.
Goldin, Lynn R.
Berrettini, Wade H.
Yoshikawa, Takeo
Sanders, Alan R.
Esterling, Lisa E.
TITLE OF INVENTION: Chromosomal Markers and Diagnostic
Tests for Manic-Depressive Illness
NUMBER OF SEQUENCES: 197
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: PastSeq for Windows Version 2.0
```

```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/091,952A
; FILING DATE: 19-Apr-1999
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/029,278
; FILING DATE: 28-OCT-1996
; APPLICATION NUMBER: PCT/US97/19381
; FILING DATE: 28-OCT-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Smith, Timothy L.
; REGISTRATION NUMBER: 35,367
; REFERENCE/DOCKET NUMBER: 015280-297100US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 277 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 1...277
; OTHER INFORMATION: rat IMP
; SEQUENCE DESCRIPTION: SEQ ID NO: 26:
US-09-091-952A-26

Query Match 11.2%; Score 193.5; DB 4; Length 277;
Best Local Similarity 25.5%; Pred. No. 8.9e-13;
Matches 70; Conservative 51; Mismatches 120; Indels 33; Gaps 10;

QY 76 ERLVEAQAADAAGEVLRKYFRQVEIIDKEDHSP---VTIADREAEAEAMVSVILKSEP 132
DB 6 QECMDYAVTLARAGEVVCIAIKNEMNMLKS---SPVDLVATDQKVEKMLISSIKEKYP 63
QY 133 THAIFGEENGWRCANSA---DFWVWLDPIDGTGKSFITGKPLFGTLIALHNGKPVIGVI 189
DB 64 SHSFIGEES-VAAGEKSILTDNPTWIIDPIDGTTFNVRFPFVAVSIGFVAVNKKIEFGVV 122
QY 190 DQPIRLERWTGVDGKQTTLNGQEISVRSNLLAQAYLYTTSPLHFEADAEDAFIRVNKV 249
DB 123 YSCVEGKMVTARKGKGAFCNGKLOVSQQEDITKSLVT---ELGSSRTPTETVRMVLNNM 179
QY 250 K----VPLYGCCDCYAVA-----LLASCFYDIVVESGLKPYDFLSLVPVIEGAGSITDWR 300
DB 180 EKLFCEIPVHGIRSVGTAANNMCLVATGGADAYYEMGHCMWDVACAGIIVTEAGGVLM DVT 239
QY 301 GDLKHPVTAESRPTSFNVVAAGD---ARVHKE 330
DB 240 GG----PFDLMSR----RVIAANNRILAERIAKE 265

RESULT 10
US-09-091-952A-28
Sequence 28, Application US/09091952A
Patent No. 6458532
GENERAL INFORMATION:
APPLICANT: Detera-Wadleigh, Sevilla D.
Gershon, Elliot S.
Badner, Judith A.
Goldin, Lynn R.
Berrettini, Wade H.
Yoshikawa, Takeo
Sanders, Alan R.
Esterling, Lisa E.
TITLE OF INVENTION: Chromosomal Markers and Diagnostic
Tests for Manic-Depressive Illness
NUMBER OF SEQUENCES: 197
CORRESPONDENCE ADDRESS:
```


Query Match 10.0%; Score 173; DB 4; Length 265;
Best Local Similarity 24.8%; Pred. No. 1.4e-10;
Matches 67; Conservative 35; Mismatches 126; Indels

Qy	91	EVLKRYQRQREIIDKEDHSP-VTIADREAEMSVILKSPPTHAIFGEE---NGWRCA 146
		: : : : : : : : : : : : : : : : : : : :
Db	9	EIIIRKAITEEKRVSTKTSAADLTETDHLVEDLIISELRERPEPSHRFIAEAAAGAKCV 68
		: : : : : : : : : : : : : : : : : : : :
Qy	147	ENSADFVWLDPDIG: {SFITGTPLFGFTLIALLNHGKPVIGIVDQPIRLRWIGVDGKT 206
		: : : : : : : : : : : : : : : : : : : :
Db	69	LTHSP-TWIIDPIDGTCNFVHRPPTVAVSIGFAVROELEFGVIYHCTEERLYTGRGRCA 127
		: : : : : : : : : : : : : : : : : : : :
Qy	207	TLNQGETSVRSNCNLLAQAALYT-----TSPHLFPEADAEDADEFIRVRNKVK 250
		: : : : : : : : : : : : : : : : : : : :
Db	128	FCNQRLRVSGETDLSKAVLVTIEIGPKRPDPATILKLFLSNMGRLLHAKAHG----- 177
		: : : : : : : : : : : : : : : : : : : :
Qy	251	VPLYGCCDYAYALLAQGFVPIVVESSGLKPYPDFLSLPVTEGAGSITDRMGDKLHPVTA 310
		: : : : : : : : : : : : : : : : : : : :
Db	178	VRVIGSTSLACLHLSGAADAYQYQULHCWDLAATAVTITREAGGITVIDTSGGGPLDMVC- 236
		: : : : : : : : : : : : : : : : : : : :
Qy	311	ESRPTSFNVVAAGD---ARVHKALDALRW 337
		: : : : : : : : : : : : : : : : : : : :
Db	237	-----RVVAASTRENMAMLIAOALOTINY 259
		: : : : : : : : : : : : : : : : : : : :

```

RESULT 14
US-09-314-198-2
; Sequence 2, Application US/09314198
; Patent No. 6403310
; GENERAL INFORMATION:
; APPLICANT: Meissner, Paul et al.
; TITLE OF INVENTION: Human Inositol Monophosphatase H1
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: MD
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:

```

```
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/314,198
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/002,072
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/US94/10465
; FILING DATE: 16-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PF134D1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 301-309-8504
; TELEFAX: 301-309-8439
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 265 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-314-198-2
```

```
Query Match 10.0%; Score 173; DB 4; Length 265;
Best Local Similarity 24.8%; Pred. No. 1.4e-10;
Matches 67; Conservative 35; Mismatches 126; Indels 42; Gaps 7;

QY 91 EVLRKYRQREVEIIDKEDHSP-VTIADREAEAMSVILKSPPTHAIFGEE---NGWRCA 146
|::|| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 9 EIIRKALTEKRVSTKTSAADLVTTETDHLVEDLIISLRERFSPHRFIAEAAAGAKCV 68
|::|| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 147 ENSADFVWVLDPRIDGPKSFITGKPLFTGLIALLHNGKPVIGVIDQPIILRERWIGVDGKQT 206
|::|| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 69 LTHSP-TWIIDPIDGTCNFVHRPPTVAVSIGFAVRQELFEGVLIYHCTEERLYTGRGRGA 127
|::|| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 207 TLNGQEISVRSCNLLAQAYLYT-----TSPHLFEADAEDAFIRVRNKVK 250
|::|| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 128 FCNGQRLRVSGETDLSKALVLTIGPKRDPATILKFLSNMGRLLHAKAHG----- 177
QY 251 VPLYGDCYAYALLASGFVDIVVESGLKPYDFLSLVPVIEGAGGSITDWRGDKLHWPVTA 310
|::|| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 178 VRVIGSSTIALCHLASGAADAYYQFGLHCWDLAAATVIRREAGGIVIDTSGGPLDLMVC- 236
|::|| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 311 ESRPTSFNVVAAGD---ARVHKEALDALRW 337
|::|| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 237 -----RVVAASTREMAMLIQAALQITINY 259
|::|| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
```

```
RESULT 15
PCT-US94-10465-3
; Sequence 3, Application PC/TUS9410465
; GENERAL INFORMATION:
; APPLICANT: M. ISSNER, ET AL.
; TITLE OF INVENTION: Human Inositol Monophosphatase H1
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
; ADDRESS: CECCHI, STEWART & OLSTEIN
; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH DISKETTE
; COMPUTER: IBM PS/2
```

```
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WORD PERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/10465
; FILING DATE: Concurrently
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: FERRARO, GREGORY D.
; REGISTRATION NUMBER: 36,134
; REFERENCE/DOCKET NUMBER: 325800-127
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 265 AMINO ACIDS
; TYPE: AMINO ACID
; STRANDEDNESS:
; TOPOLOGY: LINEAR
; MOLECULE TYPE: PROTEIN
; PCT-US94-10465-3

Query Match 10.0%; Score 173; DB 5; Length 265;
Best Local Similarity 24.8%; Pred. No. 1.4e-10;
Matches 67; Conservative 35; Mismatches 126; Indels 42; Gaps 7;

QY 91 EVLRKYRQREVEIIDKEDHSP-VTIADREAEAMSVILKSPPTHAIFGEE---NGWRCA 146
|::|| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 9 EIIRKALTEKRVSTKTSAADLVTTETDHLVEDLIISLRERFSPHRFIAEAAAGAKCV 68
|::|| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 147 ENSADFVWVLDPRIDGPKSFITGKPLFTGLIALLHNGKPVIGVIDQPIILRERWIGVDGKQT 206
|::|| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 69 LTHSP-TWIIDPIDGTCNFVHRPPTVAVSIGFAVRQELFEGVLIYHCTEERLYTGRGRGA 127
|::|| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 207 TLNGQEISVRSCNLLAQAYLYT-----TSPHLFEADAEDAFIRVRNKVK 250
|::|| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 128 FCNGQRLRVSGETDLSKALVLTIGPKRDPATILKFLSNMGRLLHAKAHG----- 177
QY 251 VPLYGDCYAYALLASGFVDIVVESGLKPYDFLSLVPVIEGAGGSITDWRGDKLHWPVTA 310
|::|| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 178 VRVIGSSTIALCHLASGAADAYYQFGLHCWDLAAATVIRREAGGIVIDTSGGPLDLMVC- 236
|::|| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 311 ESRPTSFNVVAAGD---ARVHKEALDALRW 337
|::|| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 237 -----RVVAASTREMAMLIQAALQITINY 259
|::|| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
```

Search completed: November 17, 2002, 16:43:35
Job time : 48 secs

Db 238 G 238

RESULT 2

US-09-677-064-17
; Sequence 17, Application US/09677064
; Patent No. 6291224
; GENERAL INFORMATION:
; APPLICANT: Martino-Catt, Susan J.
; APPLICANT: Wang, Hongyu
; APPLICANT: Beach, Larry R.
; TITLE OF INVENTION: Genes Controlling Phytate Metabolism in
; TITLE OF INVENTION: Plants and Uses Thereof
; FILE REFERENCE: 0706D
; CURRENT APPLICATION NUMBER: US/09/677,064
; CURRENT FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: 60/055,446
; PRIOR FILING DATE: 1997-08-11
; PRIOR APPLICATION NUMBER: 60/055,526
; PRIOR FILING DATE: 1997-08-08
; PRIOR APPLICATION NUMBER: 60/053,944
; PRIOR FILING DATE: 1997-07-28
; PRIOR APPLICATION NUMBER: 09/118,442
; PRIOR FILING DATE: 1998-07-17
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 17
; LENGTH: 267
; TYPE: PRT
; ORGANISM: Zea mays
US-09-677-064-17

Query Match 14.8%; Score 255.5; DB 4; Length 267;
Best Local Similarity 30.3%; Pred. No. 1.4e-19;
Matches 73; Conservative 41; Mismatches 112; Indels 15; Gaps 5;

QY 73 LATERLVEAQAADAAGEVLRKYFRQVEIIDKEDHSPVTIADREAEAMSVILKSFP 132
Db 1 MSEEQFLAVAEAAKSAAGEIRKGYQTKNVHKGQVDLTETDKACEDLIFNHLRHP 60
QY 133 THAIFGEENG---WRCAEASADFVWLDPIDGTGKSFITGKPLFGTLIALHNGKPVIGVI 189
Db 61 DHKFIGEEESAALCATADLTDDPTWIVDPLDGTTFNVHGFPPFCVSVGLTIGKIPTVGVV 120
QY 190 DQPIRLRERWIGVDGKQTTLNGQETSVRSCNLLAQAYLYTTSPLHFEADAEDAFIRVNKV 249
Db 121 FNPIMNELFTAVRGKGAFLNGSPKASSQDELVKALLVTEAGNRDKTTVD---DTNRI 177
250 KVPYLGCD---CYAYAL---LASGFVDIVVESGL-KPYDFLSLVPVIEGAGGSITDWR 300
Db 178 NRLLYKIRSRMCGSLALNMGVACGRDLCLCYEIGFGGPDVAAAGAVILQEGAGGLVFDPS 237
QY 301 G 301
Db 238 G 238

RESULT 3

US-09-091-952A-27
; Sequence 27, Application US/09091952A
; Patent No. 6458532
; GENERAL INFORMATION:

APPLICANT: Detera-Wadleigh, Sevilla D.
Gershon, Elliot S.
Badner, Judith A.
Goldin, Lynn R.
Berrettini, Wade H.
Yoshikawa, Takeo
Sanders, Alan R.
Esterling, Lisa E.
TITLE OF INVENTION: Chromosomal Markers and Diagnostic
Tests for Manic-Depressive Illness

NUMBER OF SEQUENCES: 197

CORRESPONDENCE ADDRESS:

ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: CA
COUNTRY: USA

ZIP: 94111-3834

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/091,952A

FILING DATE: 19-Apr-1999

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 60/029,278

FILING DATE: 28-OCT-1996

APPLICATION NUMBER: PCT/US97/19381

FILING DATE: 28-OCT-1997

ATTORNEY/AGENT INFORMATION:

NAME: Smith, Timothy L.

REGISTRATION NUMBER: 35,367

REFERENCE/DOCKET NUMBER: 015280-297100US

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 576-0200

TELEFAX: (415) 576-0300

TELEX: <Unknown>

INFORMATION FOR SEQ ID NO: 27:

SEQUENCE CHARACTERISTICS:

LENGTH: 277 amino acids

TYPE: amino acid

STRANDEDNESS: <Unknown>

TOPOLOGY: linear

MOLCULE TYPE: protein

FEATURE:

NAME/KEY: mat_peptide

LOCATION: 1...277

OTHER INFORMATION: bovine IMP

SEQUENCE DESCRIPTION: SEQ ID NO: 27:

US-09-091-952A-27

Query Match 12.2%; Score 211; DB 4; Length 277;

Best Local Similarity 26.0%; Pred. No. 1.1e-14;
Matches 67; Conservative 52; Mismatches 115; Indels 24; Gaps 8;

QY 76 ERLVEVAQRAADAAGEVLRKYFRQVEIIDKEDHSP---VTIADREAEAMSVILKSFP 132
Db 6 QECMDYAVTLGAGQGEVVRKALKNEMNIMVKS--SPADLVATDQKVEKMLITSIKEYP 63
QY 133 THAIFGEENGWRCAEASA---DFVWLDPIDGTGKSFITGKPLFGTLIALHNGKPVIGVI 189
Db 64 SHSFIGEES--VAAGEKSLTDNPTWIIDPIDGTTFNVHGFPPFVAVSFGFVYVKNKMEFGIV 122
QY 190 DQPIRLRERWIGVDGKQTTLNGQETSVRSCNLLAQAYLYTTSPLHFEADAEDAFIRVNKV 249
Db 123 YSCLEDKMYTGRKGKGAFCNGKQLVSHQEDITYKSLVT---ELGSRPTETVRIILSNI 179
QY 250 K----VPLYGCDYAYA-----LLASGFVDIVVESGLKPYDFLSLVPVIEGAGGSITDWR 300
Db 180 ERLLCPLPIHGIRGVGTAALNMLCLVAAGAADAYVEMGHCMWDVAGAGIIVTEAGGVLLDVT 239
QY 301 G---DKLHWFPVTAESRPT 315
Db 240 GGPFDLMSRRRIASSNKT 257

RESULT 4

US-09-091-952A-25
; Sequence 25, Application US/09091952A
; Patent No. 6458532

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 17, 2002, 14:18:58 ; Search time 71 Seconds
(without alignments)
634.348 Million cell updates/sec

Title: US-09-686-522c-14
Perfect score: 1730
Sequence: 1 MLSSSSVTHSDTSPFGLAS.....VVAAGDARVHKALDALRWR 338

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues
Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1730	100.0	338	21	AA442330
2	1294.5	74.8	324	21	AA442330
3	1067	61.7	295	21	AA442330
4	1020	59.0	286	23	AB93299
5	310.5	17.9	260	22	AA90637
6	310.5	17.9	260	22	AA90637
7	310.5	17.9	260	22	AA90637
8	286	16.5	265	19	AA42392
9	279	16.1	264	19	AA42388
10	267.5	15.5	271	21	AA421330

11	260	15.0	361	21	AA442330
12	260	15.0	371	21	AA442330
13	260	15.0	400	21	AA442330
14	255.5	14.8	267	20	AA97883
15	253.5	14.7	371	21	AA93265
16	253.5	14.7	400	21	AA93265
17	251.5	14.5	361	21	AA93266
18	244.5	14.1	266	22	AA67151
19	237.5	13.7	267	21	AA44227
20	236.5	13.7	233	21	AA41258
21	236.5	13.7	233	23	ABP04774
22	224	12.9	339	20	AA86325
23	221	12.8	168	21	AA44225
24	218	12.6	257	23	AB47618
25	212	12.3	249	21	AA44228
26	212	12.3	284	22	AB66131
27	210.5	12.2	273	22	AA90638
28	210.5	12.2	288	22	AB66138
29	208.5	12.1	291	22	AA979387
30	208.5	12.1	291	22	AA979388
31	208	12.0	281	23	ABP38600
32	207	12.0	254	22	AA96106
33	207	12.0	397	21	AA936298
34	204	11.8	264	22	AA55613
35	201	11.6	278	22	AB71757
36	200	11.6	284	22	AB66140
37	199.5	11.5	270	15	AA45333
38	196.5	11.4	288	19	AAW60676
39	195.5	11.3	253	22	AA82667
40	194.5	11.2	596	22	AB71755
41	193.5	11.2	277	15	AA45332
42	193.5	11.2	299	22	AB16258
43	188.5	10.9	325	20	AA935574
44	187.5	10.8	206	22	AA067657
45	185	10.7	275	23	AB32472

ALIGNMENTS

RESULT 1
AA442330

ID AA442330 standard; Protein; 338 AA.

XX AA442330;

AC AA442330;

XX 21-FEB-2000 (first entry)

DT 21-FEB-2000 (first entry)

XX Corn extragenic suppressor protein.

DE Clone cdt2c.pk003.b20; corn extragenic suppressor protein;

XX phytic acid biosynthetic enzyme; transformation; chimeric gene; antibody;

KW free phosphate; phytic acid level; transgenic plant; enzyme detection;

KW animal feed.

XX Zea mays.

OS Key

XX Location/Qualifiers

FT Misc-difference 11

FT /note= "Encoded by GCC"

FT Misc-difference 94

FT /note= "Encoded by ANG"

XX WO9955882-A1.

PN 04-NOV-1999.

XX 22-APR-1999; 99WO-US08791.

XX 24-APR-1998; 98US-0082960.

XX (DUPO) DU PONT DE NEMOURS & CO E I.

```
PI Cahoon RE, Hitz WD, Thorpe CJ, Tingey SV;
XX WPI: 2000-072181/06.
DR N-PSDB; AAZ29165.
XX
XX Novel phytic acid biosynthetic enzymes used to alter the level of the
PT enzyme in transgenic plants -
XX
XX Claim 6; Fig 2; 60pp; English.
PS
XX The present sequence is corn extragenic suppressor protein encoded by the
CC cDNA insert in clone cdt2c.pk003.b20. This belongs to the inositol
CC monophosphate family of proteins. The cDNA clone was isolated from corn
CC developing tassels. The level of expression of this phytic acid
CC biosynthetic enzyme can be altered by transforming a host cell with the
CC chimeric gene comprising the corresponding cDNA. This is done to increase
CC levels of free phosphate and decrease phytic acid levels in developing
CC seeds or in the construction of transgenic plants with altered levels of
CC the enzyme. The phytic acid biosynthetic enzymes are used to raise
CC antibodies, for enzyme detection in cells and cell extracts. This is
CC commercially used for animal feed.
XX
XX Sequence 338 AA;
Query Match 100.0%; Score 1730; DB 21; Length 338;
Best Local Similarity 100.0%; Pred. No. 1.3e-169;
Matches 338; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MLSSSSSTHSDTSPFGLASANPNRSLRLRLRAASPVSSAVLSASGROPMTVRSFAA 60
DB 1 MLSSSSSTHSDTSPFGLASANPNRSLRLRLRAASPVSSAVLSASGROPMTVRSFAA 60
QY 61 GAAGRAAAVAGELATERLVEAQAADAAGEVLRKYRQRFVEIIDKEDHSPVTIADREAE 120
DB 61 GAAGRAAAVAGELATERLVEAQAADAAGEVLRKYRQRFVEIIDKEDHSPVTIADREAE 120
QY 121 EAMVSILKSFPHTAIFGEGNWCRCENSAADFVWLDPIDGTSFTGKPLFGTLIALH 180
DB 121 EAMVSILKSFPHTAIFGEGNWCRCENSAADFVWLDPIDGTSFTGKPLFGTLIALH 180
QY 181 NGKPVIGVDOPTLRERWIGVDGKQTTLNGQEISVRSCNLLAQAYLYTTSPLHFEADAD 240
DB 181 NGKPVIGVDOPTLRERWIGVDGKQTTLNGQEISVRSCNLLAQAYLYTTSPLHFEADAD 240
QY 241 AFIRVRNKKVPLYGCDYAYALLASGFVDIVVVEGSLKPYDFLSLVPVIEGAGGSITDWR 300
DB 241 AFIRVRNKKVPLYGCDYAYALLASGFVDIVVVEGSLKPYDFLSLVPVIEGAGGSITDWR 300
/ 301 GDKLHPVTAESRPTSFNVVAAGDARVHKEALDRLWR 338
DB 301 GDKLHPVTAESRPTSFNVVAAGDARVHKEALDRLWR 338
RESULT 2
AAY44233
ID AAY44233 standard; Protein; 324 AA.
XX
XX AAY44233;
AC
XX
XX 21-FEB-2000 (first entry)
DT
XX Partial wheat extragenic suppressor protein.
DE
XX
XX Clone wim0.pk0010.f6; wheat extragenic suppressor protein;
KW phytic acid biosynthetic enzyme; transformation; chimeric gene; antibody;
KW free phosphate; phytic acid level; transgenic plant; enzyme detection;
KW animal feed.
XX
XX Triticum aestivum.
OS
XX
XX WO955882-A1.
PN
XX
XX 04-NOV-1999.
PD
```

```
XX 22-APR-1999; 99WO-US08791.
XX
XX 24-APR-1998; 98US-0082960.
XX
XX (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
XX Cahoon RE, Hitz WD, Thorpe CJ, Tingey SV;
PI
XX WPI: 2000-072181/06.
DR N-PSDB; AAZ29168.
XX
XX Novel phytic acid biosynthetic enzymes used to alter the level of the
PT enzyme in transgenic plants -
XX
XX Claim 6; Fig 2; 60pp; English.
PS
XX The present sequence is a portion of wheat extragenic suppressor protein
CC encoded by the cDNA insert in clone wim0.pk0010.f6. This belongs to the
CC inositol monophosphate family of proteins. The cDNA clone was isolated
CC from wheat seedlings. The level of expression of this phytic acid
CC biosynthetic enzyme can be altered by transforming a host cell with the
CC chimeric gene comprising the corresponding cDNA. This is done to increase
CC levels of free phosphate and decrease phytic acid levels in developing
CC seeds or in the construction of transgenic plants with altered levels of
CC the enzyme. The phytic acid biosynthetic enzymes are used to raise
CC antibodies, for enzyme detection in cells and cell extracts. This is
CC commercially used for animal feed.
XX
XX Sequence 324 AA;
Query Match 74.8%; Score 1294.5; DB 21; Length 324;
Best Local Similarity 80.6%; Pred. No. 9.8e-125;
Matches 254; Conservative 24; Mismatches 28; Indels 9; Gaps 3;
QY 32 LRAASPVSSAVLSASGR-----OPMSTVRASFAAGRAAAGVGE--LATERLVEAQ 83
DB 11 LRPSGLIATFSSAAGRACGTAGRMGVSVRAS-PSEAGGWAFAAAGKEGVDMERLYAVAQ 69
QY 84 RAADAAGEVLRKYRKYR VVEIIDKEDHSPVTIADREAEAEAMVSILKSFPHTAIFGEGN 143
DB 70 SAADAAGEVLRKYRQRFVEIIDKEDHSPVTIADREAEAEAMVSILKSFPHTAIFGEGN 129
QY 144 RCAENSADFVWLDPIDGTSFTGKPLFGTLIALHNGKPVIGVIDOPILRERWIGVDG 203
DB 130 RCAEKSADYVWLDPIDGTSFTGKPLFGTLIALHNGKPVGMGIIDOPILRERWVGVDG 189
QY 204 KQTTLNGQEISVRSCNLLAQAYLYTTSPLHFEADAEAFIRVRNKKVPLVYGCDCYAYAL 263
DB 190 KQTTLNGQEISVRPCNVLEQAYLYTTSPLHFEADAEAFIRVRDKVPLVYGCDCYAYAL 249
QY 264 LASGFVDIVVVEGSLKPYDFLSLVPVIEGAGGSITDWRGDKLHPVTAESRPTSFNVVAAG 323
DB 250 LASGFVDIVVVEGSLKPYDFLSLVPVIEGAGGSITDMEGNKLNHPVSSSRPTSFNVVAAG 309
QY 324 DARVHKEALDRLWR 338
DB 310 DSHVHGQALALRLWR 324
RESULT 3
AAY44232
ID AAY44232 standard; Protein; 295 AA.
XX
XX AAY44232;
AC
XX 21-FEB-2000 (first entry)
DT
XX Soybean extragenic suppressor protein.
DE
XX
XX Clone s12.pk122.p24; clone src3c.pk013.g15; clone sflin.pk003.g19;
KW soybean extragenic suppressor protein; phytic acid biosynthetic enzyme;
KW transformation; chimeric gene; antibody; free phosphate;
KW
```


CC metabolism and oxidative phosphorylation (SMP) proteins given in
 CC AAB79243 to AAB 79633 which are involved in carbon metabolism and
 CC energy production. The C. glutamicum SMP gene can be used in vectors
 CC (II) for expression in host cells and production or modulation of
 CC production of fine chemicals, such as, an organic acid, a proteinogenic
 CC or nonproteinogenic amino acid (preferred), a purine or pyrimidine base,
 CC a nucleoside, a nucleotide, a lipid, a saturated or unsaturated fatty
 CC acid, a diol, a carbohydrate, an aromatic compound, a vitamin, a
 CC cofactor, a polyketide, or an enzyme. The presence of (I) or SMP proteins
 CC (III) encoded by them are used for diagnosing the presence or activity of
 CC Corynebacterium diphtheriae in a subject. (I), (II), (III) or host cells
 CC containing them are used to map genomes of organisms related to
 CC C. glutamicum, identify and localise C. glutamicum sequences of interest,
 CC in evolutionary studies, in determining SMP protein regions required
 CC for function, in modulating SMP protein activity, in modulating the
 CC metabolism of sugars, and in modulating high-energy molecule production
 CC in a cell (i.e. ATP, NADPH).
 XX
 SQ Sequence 260 AA;
 Query Match 17.9%; Score 310.5; DB 22; Length 260;
 Best Local Similarity 33.3%; Pred. No. 2.2e-23;
 Matches 91; Conservative 42; Mismatches 101; Indels 39; Gaps 11;
 QY 81 VAQRAADAAGEV-LRKYPQRQVEIIDKEDHSPVTIADREAEAMVSVILKSFPTFAIFGE 139
 DB 10 LALELAELADSTLDRFEASDLVSSKPDMPVSDADLATEEALREKIATARPADSILGE 69
 QY 140 ENGWRCNAENSADF---VWVLDPIDGTSFKITGKPLFGFLIALLHNGKP GVIDQPIILRE 196
 DB 70 EFG-----GDVEFSQRWIIIDIDGKNYVRGVPWWTALIALLDNGKPVAGVISAPALAR 124
 QY 197 RWIGVIGKQT--TLNG---QEISVRSCNLLAQAYL-YTTPHLPFEADAEDAFIRVRNKV- 249
 DB 125 RWWASEGAGAMRTFNGSPKRLSVSOVSKLDDASLSFSSLSGWAERDLRDQFVSLTDITW 184
 QY 250 KVPILGCDYAYALLASGFVDIVVESGLKPYDFLSL-VIEGAGGSTLDWRG-DKLHWPV 308
 DB 185 RLRGYG-DFFSYCLVAEGADVIAAEPEVSLWDLAPLSILVTEAGGKFTSLAGVDGPH--- 240
 QY 309 TAESRPTSFNVVAAGDA-----RVHKEALDALR 336
 DB 241 -----GGDAVATNGILHDETLDRLK 260
 RESULT 7
 AAB79382
 ID AAB79382 standard; Protein: 260 AA.
 IX AAB79382;
 AC AAB79382;
 XX 30-APR-2001 (first entry)
 XX Corynebacterium glutamicum SMP protein sequence SEQ ID NO:280.
 DE Corynebacterium glutamicum; carbon metabolism and energy production;
 KW SMP protein; sugar metabolism and oxidative phosphorylation protein;
 KW fine chemical production; organic acid; proteinogenic amino acid;
 KW nonproteinogenic amino acid; purine base; pyrimidine base; nucleoside;
 KW nucleotide; lipid; saturated fatty acid; unsaturated fatty acid; diol;
 KW carbohydrate; aromatic compound; vitamin; cofactor; polyketide; enzyme;
 KW diagnosis; Corynebacterium diphtheriae; evolutionary study.
 XX
 OS Corynebacterium glutamicum.
 XX
 PN WO200100844-A2.
 XX
 XX 04-JAN-2001.
 PD
 XX 23-JUN-2000; 2000WO-IB00943.
 PF
 XX 25-JUN-1999; 99US-0141031.
 PR 08-JUL-1999; 99DE-1031412.

PR 08-JUL-1999; 99DE-1031413.
 PR 08-JUL-1999; 99DE-1031419.
 PR 08-JUL-1999; 99DE-1031420.
 PR 08-JUL-1999; 99DE-1031424.
 PR 08-JUL-1999; 99DE-1031428.
 PR 08-JUL-1999; 99DE-1031431.
 PR 08-JUL-1999; 99DE-1031433.
 PR 08-JUL-1999; 99DE-1031434.
 PR 08-JUL-1999; 99DE-1031510.
 PR 08-JUL-1999; 99DE-1031562.
 PR 08-JUL-1999; 99DE-1031634.
 PR 09-JUL-1999; 99DE-1032180.
 PR 09-JUL-1999; 99DE-1032227.
 PR 09-JUL-1999; 99DE-1032230.
 PR 09-JUL-1999; 99US-0143208.
 PR 14-JUL-1999; 99DE-1032424.
 PR 14-JUL-1999; 99DE-1032973.
 PR 14-JUL-1999; 99DE-1033005.
 PR 27-AUG-1999; 99DE-1040765.
 PR 31-AUG-1999; 99US-0151572.
 PR 03-SEP-1999; 99DE-1042076.
 PR 03-SEP-1999; 99DE-1042079.
 PR 03-SEP-1999; 99DE-1042086.
 PR 03-SEP-1999; 99DE-1042087.
 PR 03-SEP-1999; 99DE-1042088.
 PR 03-SEP-1999; 99DE-1042095.
 PR 03-SEP-1999; 99DE-1042123.
 PR 03-SEP-1999; 99DE-1042125.
 XX (BADI) BASF AG.
 XX Pompejus M, Kroegeer B, Schroeder H, Zeider O, Haberhauer G;
 PI WPI; 2001-061975/07.
 XX N-PSDB; AAF71499.
 DR
 DR
 XX
 PT New isolated Corynebacterium glutamicum nucleic acid encoding a sugar
 PT metabolism and oxidative phosphorylation protein for production or
 PT modulation of production of fine chemicals e.g. amino acids,
 PT carbohydrates or enzymes -
 XX
 PS Claim 20; Page 540; 1246pp; English.
 XX
 XX AAF71360 to AAF71750 encode the Corynebacterium glutamicum sugar
 CC metabolism and oxidative phosphorylation (SMP) proteins given in
 CC AAB79243 to AAB 79633 which are involved in carbon metabolism and
 CC energy production. The C. glutamicum SMP gene can be used in vectors
 CC (II) for expression in host cells and production or modulation of
 CC production of fine chemicals, such as, an organic acid, a proteinogenic
 CC or nonproteinogenic amino acid (preferred), a purine or pyrimidine base,
 CC a nucleoside, a nucleotide, a lipid, a saturated or unsaturated fatty
 CC acid, a diol, a carbohydrate, an aromatic compound, a vitamin, a
 CC cofactor, a polyketide, or an enzyme. The presence of (I) or SMP proteins
 CC (III) encoded by them are used for diagnosing the presence or activity of
 CC Corynebacterium diphtheriae in a subject. (I), (II), (III) or host cells
 CC containing them are used to map genomes of organisms related to
 CC C. glutamicum, identify and localise C. glutamicum sequences of interest,
 CC in evolutionary studies, in determining SMP protein regions required
 CC for function, in modulating SMP protein activity, in modulating the
 CC metabolism of sugars, and in modulating high-energy molecule production
 CC in a cell (i.e. ATP, NADPH).
 XX
 SQ Sequence 260 AA;
 Query Match 17.9%; Score 310.5; DB 22; Length 260;
 Best Local Similarity 33.3%; Pred. No. 2.2e-23;
 Matches 91; Conservative 42; Mismatches 101; Indels 39; Gaps 11;
 QY 81 VAQRAADAAGEV-LRKYPQRQVEIIDKEDHSPVTIADREAEAMVSVILKSFPTFAIFGE 139
 DB 10 LALELAELADSTLDRFEASDLVSSKPDMPVSDADLATEEALREKIATARPADSILGE 69
 QY 140 ENGWRCNAENSADF---VWVLDPIDGTSFKITGKPLFGFLIALLHNGKPVICVIDOPIILRE 196

Db 70 EFG-----GDVEFSGROWIIDPIDGKKNYVRGYPVWATLIALLDNGKPVAGVISAPALAR 124
QY 197 RWIGVDGKQT--TLNG---QEISVRSCNLLAQAYL-YTSPHLPFEADAEDAFIRVRNKV- 249
Db 125 RWASEGAGAWRTFNGSPKLSVSQVSKLDDASLSFSLSGWAERDLRDOFVSLTDTTW 184
QY 250 KVPYGCDCYAYALLASGFVDIVVESGLKPYDFLSLPVPVIEGAGGSITDWRG-DKLRHPV 308
Db 185 RLKGYG-DFFSYCLVAEGAVDIAAEPEVSLWDLAPLSILVTEAGGKFTSLAGVDGPH--- 240
QY 309 TABSRPTSFNVVAAGDA-----RVHREALDALR 336
Db 241 -----GGDAVATNGILHDETLDRK 260

RESULT 8
AAW42392
ID AAW42392 standard; Protein; 265 AA.
XX AAW42392;
DT 22-JUN-1998 (first entry)
XX Aquifex pyrophilus phosphatase (28ph1).
DE Alkaline phosphatase; thermostable enzyme; thermophilic bacterium;
KW food; detergent; baking.
XX Aquifex pyrophilus.

OS
XX WO9748416-A1.
XX 24-DEC-1997.
XX 19-JUN-1997; 97WO-US10784.
XX 19-JUN-1996; 96US-0033752.
XX (RECO-) RECOMBINANT BIOCATALYSIS INC.
XX Bylina E, Lee E, Mathur EJ;
PI WPI: 1998-062851/06.
DR N-PSDB; AAV03317.
XX Thermostable phosphatase(s) - useful in pharmaceutical, food,
PT detergent, and baking industries
PS Claim 11; Page 95; 128pp; English.

XX This protein comprises a thermostable phosphatase, designated
CC 28ph1, of Aquifex pyrophilus. The invention relates to claimed
CC polynucleotides (see AAV03301-20) coding for claimed thermostable
CC phosphatases (see AAW42380-95). Vector and host cells are used to
CC produce the enzymes, which can be used in a claimed method to
CC hydrolyse phosphate bonds. They can also be used in enzyme
CC labelling processes, in certain recombinant DNA techniques, in
CC ELISA immunoassays, in enzyme linked gene probes, in research
CC applications for removing 5' phosphates in polynucleotides prior to
CC end labelling, and in the pharmaceutical, food, detergent, and
CC baking industries.

XX Sequence 265 AA;
Query Match 16.5%; Score 286; DB 19; Length 265;
Best Local Similarity 29.8%; Pred. No. 7.7e-21;
Matches 78; Conservative 53; Mismatches 105; Indels 26; Gaps 8;

QY 76 ERLVEVAQRAADAGEVLRKYF-RQVE-IIDKEDHSPVTIADREAEAMVSVILKSFT 133
Db 5 EKLLEVAKMAALAGGQVLKFNFKIKLENTEKGEKDFVSVDKTSSEERIKELILKFFPD 64

QY 134 HAIFGEENWRCAESADFVWVLDPIDGTSFTGKPLFGLTIALLLHNGKPVIGVIDOPI 193
Db 65 HEVVGEEKEKE--SPYKWFIDPLDGTKNYIKGFIFAVSVGLKENEPIVAGVILPY 122
QY 194 LRRWTIGVDCQKQTTLNGQEISVRSCNLLAQAYLYTTSPLFEADA-----EDAFIRVR 246
Db 123 FDTLYWASKRGAYKNGERISVKERGELKHAADVVGFPSSRRDISLYLNVFKEVFEV- 181
QY 247 NKVKVPLYGCDCYAYALLASGFVDIVVESGLKPYDFLSLPVPVIEGAGGSITDWRGDKLHW 306
Db 182 GSVRRP--GAAAVDICIMLAEGIFDGMFEPMKPDITAGLVILKEAGGYT-LKGD---- 234
QY 307 PVTASRPTSFNVVAAGDARVH 328
Db 235 -----PFGISDIAGNRMLH 249

RESULT 9
AAW42388
ID AAW42388 standard; Protein; 264 AA.
XX AAW42388;
DT 22-JUN-1998 (first entry)
XX Aquifex VF-5 alkaline phosphatase (34Ala).
DE Alkaline phosphatase; thermostable enzyme; thermophilic bacterium;
KW food; detergent; baking.
XX Aquifex sp. strain VF-5.

OS
XX WO9748416-A1.
XX 24-DEC-1997.
XX 19-JUN-1997; 97WO-US10784.
XX 19-JUN-1996; 96US-0033752.
XX (RECO-) RECOMBINANT BIOCATALYSIS INC.
XX Bylina E, Lee E, Mathur EJ;
PI WPI: 1998-062851/06.
DR N-PSDB; AAV03309.
XX Thermostable phosphatase(s) - useful in pharmaceutical, food,
PT detergent, and baking industries
PS Claim 10; Fig 9; 128pp; English.

XX This protein comprises a mature thermostable alkaline phosphatase
CC (34Ala) of Aquifex VF5, a marine bacterium that grows optimally
CC at 85-90 degC and pH 6.8. Its amino acid sequence, deduced from
CC an isolated polynucleotide (see AAV03309), shows 57% identity to
CC Escherichia coli suppressor protein subB. The invention relates to
CC claimed polynucleotides (see AAV03301-09 and AAV03310-20) coding for
CC claimed thermostable phosphatases (see AAW42380-95). Vector and host
CC cells are used to produce the enzymes, which can be used in a
CC claimed method to hydrolyse phosphate bonds. They can also be used
CC in enzyme labelling processes, in recombinant DNA techniques, in
CC ELISA, in enzyme linked gene probes, in research applications for
CC removing 5' phosphates in polynucleotides prior to end labelling,
CC and in the pharmaceutical, food, detergent, and baking industries.

XX Sequence 264 AA;
Query Match 16.1%; Score 279; DB 19; Length 264;
Best Local Similarity 28.5%; Pred. No. 4e-20;
Matches 75; Conservative 54; Mismatches 106; Indels 28; Gaps 7;

QY 76 ERLVEVAQRAADAGEVLRKYF--QVEIIDKEDHSPVTIADREAEAMVSVILKSFT 133

```

Db      5  KYLEVAKTAAAGQVLFKFKVKKKEEKEGDFVSYVDKTSERIKKEVILKFFPD 64
QY     134 HAIFGEENGWRCAES-ADFVWVLDPIDGTSKSPITGKPLFGTLIALHNGKPVIGVIDQP 192
Db      65 HEVVGEMG--AEGSGSEYRWFIDPLDGTCKNYINGFPPIFAVSVGLVKGEPIVGAVILP 121
QY     193 ILRERWIGVDGKQTTLNGQSEISVRSNLLAQAYLYTTSPLFEADA-----EDAFIRV 245
Db     122 YFDKLMGAKGLGAYNGRKIKVDNLSLKHAGVYGFPSRRRDISIYLNIFKQVFEV 181
QY     246 RNKVKPLYCCDCYAVALLASGPDVWESGLKPYDFLSLPVIEGAGGSITDWRGDKLH 305
Db     182 -GSMRRP--GAAAVDLCHVAEGIFDGMFEEMKPDITAGLVILKEAGG----- 227
QY     306 WPVTAESRPTSFNVNAGDARVH 328
Db     228 -VYTLGCEPFGVSDIITAGNKAH 249

RESULT 10
AAG21330
ID AAG21330 standard; Protein; 271 AA.
XX
AC AAG21330;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 23850.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144425.
PR 19-JUL-1999; 99US-0144431.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
```


PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135114.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158222.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.

[illegible]

Oy 301 G 301 PR 18-JUN-1999; 99US-0139454.
| 18-JUN-1999; 99US-0139455.
Db 238 G 238 PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 18-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.

RESULT 15
AAG39265
ID AAG39265 standard; Protein; 371 AA.
XX AC AAG39265;
XX DT 18-OCT-2000 (first entry)
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 48558.
XX DE Arabidopsis thaliana.
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX Arabidopsis thaliana.
N EPI033405-A2.
XX PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-0301439.
XX PF 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127482.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 17, 2002, 15:16:18 ; Search time 85 Seconds
(without alignments)
819.340 Million cell updates

Title: US-09-686-522C-14
 Perfect score: 1730
 Sequence: 1 MLSSSSSTHSDTSPFPGLAS.....VVAAGDARVHKEALDALRWR 338

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

total number of hits satisfying chosen parameters: 671580

```
Minimum DB seq length: 0
```

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 1008

Listing first 45 summaries

Database :

```

SPFEMBL_21.*
1:  sp.archae.*
2:  sp.bacteria.*
3:  sp.fungi.*
4:  sp.human.*
5:  sp.invertebrate.*
6:  sp.mammal.*
7:  sp.mhc.*
8:  sp.organelle.*
9:  sp.phase.*
10: sp.plant.*
11: sp.protein.*
12: sp.virus.*
13: sp.vertebrate.*
14: sp.unclassified.*
15: sp.rvirus.*
16: sp.bacteria.*
17: sp.archae.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	1020	59.0		286	10	Q9T021	arabidopsis
2	551	31.8		259	16	Q9N311	rhizobium m
3	525	30.3		297	16	Q8U945	agrobacteri
4	483	27.9		308	16	Q930E4	rhizobium m
5	477	27.6		278	16	Q9A643	caulobacter
6	463.5	26.8		272	16	Q8YDX6	brucella me
7	460	26.6		263	2	Q9KNB5	agrobacteri
8	449	26.0		262	16	Q9NH43	rhizobium l
9	396.5	22.9		267	16	P73806	synecocyst
10	317	18.3		239	16	Q988Q3	rhizobium l
11	317	18.3		273	2	O53743	streptomyce
12	302.5	17.5		268	2	Q8RJW8	streptomyce
13	293	16.9		267	16	Q8ZCR8	yersinia pe
14	283	16.4		255	16	O32889	mycobacteri
15	278.5	16.1		266	16	Q9K4B1	streptomyce
16	273.5	15.8		260	16	P95189	mycobacteri

ALIGNMENTS

```

RESULT 1
Q9T021
ID Q9T021 PRELIMINARY; PRT: 286 AA.
AC Q9T021;
DT 01-MAY-2000 (TrEMBLre 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE Inositol monophosphatase-like protein.
GN T22F8.20 OR AT4G39120.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID:3702;
[1]
SEQUENCE FROM N.A.
RA Bevan M., Zimmermann W., Grueniseisen A., Wambutt R., Bancroft I.,
RA Mewes H.W., Mayer K.F.X., Schueller C.;
RA Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
[2]
SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RA Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
[3]
SEQUENCE FROM N.A.
RA Zimmermann W., Grueniseisen A., Wambutt R., Kalicki J., Wohldmann P.,
RA Smith A., Mewes H.W., Lemcke K., Mayer K.F.X.;
RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
[4]
SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL050351; CAB43627.1; -
DR EMBL: AL161594; CAB80575.1; -
DR InterPro: IPR000760; Inositol_P.
DR Pfam: PF00459; inositol_P; 1.
DR PROSITE; PS00629; IMP_1; 1.
SQ SEQUENCE 286 AA; 31635 MW; 5C2BB2BA90FD10BB CRC64;

```


DE	Inositol monophosphatase family protein.
GN	CC2252.
OC	Caulobacter crescentus.
OC	Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OC	Caulobacter.
OX	NCBI_TaxID=155892;
OX	[1]
RN	SEQUENCE FROM N.A.
RP	STRAIN=ATCC 19089 / CB15;
RC	MEDLINE=21173698 / PubMed=11259647;
RC	Nierman W.C., Feldbylum T.V., Laub M.R.K., Ohta N., Maddock J.R.,
RA	Eisen J., Heideberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA	Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA	DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA	Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA	utterback T., Tran K., Wolf A., Yamathavan J., Ermolaeva M., White O.,
RA	Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT	*Complete genome sequence of Caulobacter crescentus.*;
RL	Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR	EMBL; AF005895; AAK24223.1; -;
DR	TIGR; CC2252;
DR	InterPro; IPRO00760; Inositol_P.
DR	Pfam; PF00459; inositol_P; 1.
KW	Complete proteome.
QY	SEQUENCE 278 AA; 29688 MW; 281299D0F028757A CRC64;
QY	Query Match 27.6%; Score 477; DB 16; Length 278;
QY	Best Local Similarity 40.8%; Pred. No. 1.6e-30;
QY	Matches 116; Conservative 48; Mismatches 88; Indels 32; Gaps
QY	73 LATERLVEVAQRAAD---AAGEVLRYFRQRVEIIDK-----EDHS---PVTIADRE 118
QY	: :
Db	3 LSADRLAAALDAFDLLDNRASADVILPLFRADHGLDKGAGKNLPDTHAAAFDPVTEADR 62
QY	119 AESAMSVILKSPTTIAIFGEENGWRCAEN--SADFVWVLDPIDGTKSPITCKPLFGTLI 176
QY	: :
Db	63 AEAAIRALTAQRPDDHGVICEEYG----EDRPDAEFVWLDPIDGTGRAFIAGLPLWTTLI 118
QY	177 ALLHNKPVGIVTDQIPILRWGITGVCGKQTTLNGQE---ISVRSNCLLAQAAYLYTTSPIH- 232
QY	: :
Db	119 GLRHGEPPVLGSICQPVNFEIFIGHAGARLVSGGEARPVRVCANINDAVIATDPDA 178
QY	233 LFPRADAEDAFIRVNKKVPYLGCDCCYAYALLASGEVDIVWSGLKPYDFLSLVPVIEGA 292
QY	: :
Db	179 CFQDGERGAWQLVRAAKLARLGCDAYAYAMVAMGMKMVIEAGLSWDIEAIIPIEGA 238
QY	293 GGSIITDWRGDKLHPVTAESRSPTSFNVAAGDARVHKELDRLR 336
QY	: :
Db	239 GGWTVNRGR-----PW---GPNGGGMVISGDHRPLDEALVSRLR 274
RESULT 6	
ID	OBYDX6 PRELIMINARY; PRT; 272 AA.
ID	OBYDX6
ID	OBYDX6
DT	01-MAR-2002 (TEMBLrel. 20, Created)
DT	01-MAR-2002 (TEMBLrel. 20, Last sequence update)
DT	01-JUN-2002 (TEMBLrel. 21, Last annotation update)
DE	Extragenic suppressor protein SUHB.
OS	BMEI10048.
OS	Brucella melitensis.
OC	Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC	Brucellaceae; Brucella.
OX	NCBI_TaxID=29459;
OX	[1]
RN	SEQUENCE FROM N.A.
RP	STRAIN=16M / ATCC 23456 / BIOTYPE 1;
RC	MEDLINE=2002109; PubMed=11756688;
RC	DelVecchio V.G., Kapral V., Redkar R.J., Patra G., Mujer C., Los T.
RA	Ivanova N., Anderson I., Bhattacharya A., Lykidis A., Reznik G.,
RA	Jablonski L., Larsen N., D'Souza M., Bernal A., Mazur M., Goltsman E.,
RA	Salkov E., Elzer P.H., Hagius S., O'Callaghan D., Letesson J.-J.,
RA	Haskell-Korn R., Kyripides N., Overbeek R.;

RT *The genome sequence of the facultative intracellular pathogen
 RT Brucella melitensis.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:443-448(2002).
 DR EMBL; AA009643; AAL53289.1; -
 DR InterPro; IPR000760; Inositol_P.
 DR InterPro; IPR000146; IO_FB_phptase.
 DR Pfam; PF00459; Inositol_P; 1.
 DR PRINTS; PR00377; INFPHPTASE.
 DR PROSITE; PS00629; IMP_1; 1.
 KW Complete proteome.
 SQ SEQUENCE 272 AA; 29222 MW; 42C32EAF608C7E26 CRC64;

Query Match 26.8%; Score 463.5; DB 16; Length 272;
 Best Local Similarity 41.3%; Pred. No. 1.9e-29;
 Matches 116; Conservative 47; Mismatches 93; Indels 25; Gaps 9;

QY 66 RAAAVGELATERLEVA---QRAADAAGEVLKRYRQRVEIIDKED--HSPVTIADREAE 120
 DB 4 RAPCFGNRTGLLDKAFSEVAAAAAQTLPF-FRQUTEVDNKKYSVGFDPVTEADRAE 62
 ' 121 EAMSVILKSPFTHAIFGEENGWRCNAENS-ADENVVWLDPIDGKTSFTGKPLFGTLIAL 179
 DB 63 RAIRAVIGETFPYHGLGEEYG---AENTDRSHWIIIPVDGTRAFISGLPWGTLVGT 119
 QY 180 HNGKPVIGVIDQPIILRRERWIGVDGKOTTLN-----QELSVRSNLLAQAYLYTTSPLHF 234
 DB 120 VDCDARAGMSPFTGELFYS-DGCGAVLQRGDGPAPRLCVRKNAVLDEATLTFTTTPALF 178
 QY 235 EADAEDAFIRVNKVKVPLVGCDCYAYALLASGFFVDIVVESGLKPYDFLSLVPIYEGAGG 294
 DB 179 KGDDRKAFDLRESARLSRYSVDCYAFAMLAGGFVDIVVEAGLQTYDIAALIPIEQAGG 238
 QY 295 SITDRGDKLHPVPTAESRPTSFNVVAAGDARVHKEALDAL 335
 DB 239 VTTRDGG-----PAEQ---CGDIVAATPALHQAALDLL 270

RESULT 7

ID Q9KWB5 PRELIMINARY; PRT; 263 AA.
 AC Q9KWB5;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE Riorf69 protein.
 GN RIORF69.
 OS Agrobacterium rhizogenes.
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OX NCBI_TaxID=359;
 JX NCBI_TaxID=359;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-MAFF03-01724; PLASMID-PR11724;
 RA Moriguchi K., Maeda Y., Satou M., Kataoka M., Tanaka N., Yoshida K.;
 RT "Analysis of unique variable region of a plant root inducing plasmid,
 RT pr11724, by the construction of its physical map and library.";
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-MAFF03-01724; PLASMID-PR11724;
 RA Moriguchi K., Nishida T., Maeda Y., Tanaka N., Yoshida K.;
 RT "Genome structure of Ri plasmid (1): Construction of linking library
 RT and physical map of Ri1724 in Japanese Agrobacterium.";
 RL Nucleic Acids Symp. Ser. 39:189-190(1998).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-MAFF03-01724; PLASMID-PR11724;
 RA Moriguchi K., Maeda Y., Satou M., Satuti N., Tanaka N.,
 RA Yoshida K.;
 RT "The complete nucleotide sequence of a Ri (root inducing) plasmid
 RT indicates its chimerical structure between Ti and Sym plasmids.";
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.

RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN-MAFF03-01724; PLASMID-PR11724;
 RA Moriguchi K., Maeda Y., Satou M., Kataoka M., Tanaka N., Yoshida K.;
 RT "Analysis of unique variable region of a plant root inducing plasmid,
 RT pr11724, by the construction of its physical map and library.";
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN-MAFF03-01724; PLASMID-PR11724;
 RA MEDLINE-20241294; PubMed-10780382;
 RA Maeda Y., Moriguchi K., Kataoka M., Satou M., Satuti N., Tanaka N.,
 RA Yoshida K.;
 RT "Genome structure of Ri plasmid (1): Sequencing analysis of T-DNA and
 RT its flanking regions of pr11724 in Japanese Agrobacterium
 RT rhizogenes.";
 RL Nucleic Acids Symp. Ser. 42:67-68(1999).
 DR EMBL; AB039932; BAA97780.1; -
 DR EMBL; AP002086; BAB16188.1; -
 DR HSSP; P29218; IIMF.
 DR InterPro; IPR000760; Inositol_P.
 DR Pfam; PF00459; Inositol_P; 1.
 DR PROSITE; PS00629; IMP_1; 1.
 KW Plasmid.
 SQ SEQUENCE 263 AA; 28110 MW; 2979EAF9F8958A25F CRC64;

Query Match 26.6%; Score 460; DB 2; Length 263;
 Best Local Similarity 38.8%; Pred. No. 3.4e-29;
 Matches 107; Conservative 55; Mismatches 82; Indels 32; Gaps 8;

QY 74 ATERLVE---VAQRAADAAGEVLKRYRQRVEIIDK--EDHSPVTIADREAEAMVSYL 128
 DB 3 STEDVISDFELRLAAVAARETLPR-FRLPGTVVNKIAGGDFPVTADRETEKALRALIR 61
 QY 129 KSFTHAIFGSENGWRCNAENSADFVWVLDPIDGKTSFTGKPLFGTLTALLHNGKPVIGV 188
 DB 62 SEYPGHGVLGEEYGNEGAD--AELVWVIDPIDGTRSFISGLPGLVGLTRDGDVAGM 119
 QY 189 IDQPIILRRERWIGVDGKOTTLNG-----QELSVRSNLLAQAYLYTTSPLHFEADAE 239
 DB 120 MAQPFIFELF-----YSTGNGAFYEGPHGOTOLATSATKDLADATLTFTTTPAFEDRR 173
 QY 240 DAFIRVNKVKVPLVGCDCYAYALLASGFFVDIVVESGLKPYDFLSLVPIYEGAGGSITDM 299
 DB 174 SSYDILEKQVRLARYGTDYCYACMLAAGIDCVVEAGLYPYDIALIPIEQAGGVVTDW 233
 QY 300 RGDKLHPVPTAESRPTSFNVVAAGDARVHKEALDAL 335
 DB 234 DGG-----PAE---KCGGI VAAANATIIHAQIMACL 260

RESULT 8

ID Q98H43 PRELIMINARY; PRT; 262 AA.
 AC Q98H43;
 DT 01-OCT-2001 (TrEMBLrel. 18, Created)
 DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Mono-phosphatase.
 GN MLL3039.
 OS Rhizobium loti (Mesorhizobium loti).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OX NCBI_TaxID=381;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-MAFF303099;
 RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
 RA Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,
 RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
 RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,
 RA Takeuchi C., Yamada M., Tabata S.;

```
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RL Mesorhizobium loti."
RL DNA Res. 7:331-338(2000).
DR EMBL: AF003001; BAB50023.1; -.
DR InterPro: IPR000760; Inositol_P.
DR Pfam: PF00459; inositol_P; 1.
DR PROSITE: PS00629; IMP_1; 1.
KW Complete proteome.
SQ SEQUENCE 262 AA; 27882 MW; E38C73FB36D27CE5 CRC64;

Query Match 26.0%; Score 449; DB 16; Length 262;
Best Local Similarity 39.8%; Pred. No. 2.6e-28;
Matches 105; Conservative 50; Mismatches 91; Indels 18; Gaps 7;

QY 79 VEVAQRADAAGEVLKRYRQREVEIIDKEDHS--PVTIADRAEAEAMSVILKSPPTTHAI 136
Db 10 IDFMRIQAQAETAETPRFAQGAANKKSGFDPVTEADRETAIRALIAQYPDHGI 69

QY 137 FGEENGWRCAN--SADFVWVLDPIDGTSFTGKPLFGLTIALLLHNGKPVIGVIDOPILR 195
b 70 LGEHGG---SENISRRHWVIDPIDGTRAFISGLPVWGTLVGLTVDGDAVAGMMAQPFETG 126

QY 196 ER-WIGVDGKQTTLNG--QEISVRSNLLAQAYLYTTSPLHFEADAEDAFIRVKNKVYP 252
Db 127 ELFYANASGSHYEGPGPKRLSKTRTLDEATLFTTTFALFPGEARTRYDAFEKOVOLA 186

QY 253 LGGDCYAYALLASGFVDIVVESGLKPYDFLSLPVPVIEGAGSGITDWRGDKLHWPVTAES 312
Db 187 RYGCADYAFAMIASGSDIVADPGLKPYDIVALPIIERAGGVVTFDGG-----PAEK 240

QY 313 RPTSFNVVAAGDARVHKEALDALR 336
Db 241 ---GGDVLAAATPELHAAAMAALR 261

RESULT 9
P73806 PRELIMINARY; PRT; 267 AA

AC P73806;
DT 01-FEB-1997 (TREMBLrel. 02, Created)
DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Extragenic suppressor.
GN SUH8 OR SLL1959.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97061201; PubMed=8905231;
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hirose M., Sugita M., Saito M., Saito T., Kimura T.,
RA Hosouchi T., Matsumoto A., Muraki A., Nakazaki N., Naruo K., Okumura S.,
RA Shimpo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
RA Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-136(1996).
DR EMBL: D90909; BAA17860.1; -.
DR InterPro: IPR000760; Inositol_P.
DR Pfam: PF00459; inositol_P; 1.
DR PROSITE: PS00629; IMP_1; 1.
KW Complete proteome.
SQ SEQUENCE 267 AA; 29243 MW; 3E0A3536C64AE707 CRC64;

Query Match 22.9%; Score 396.5; DB 16; Length 267;
Best Local Similarity 36.9%; Pred. No. 4.7e-24;
Matches 101; Conservative 49; Mismatches 99; Indels 25; Gaps 10;

QY 76 ERLVEVQRAADAAGEVLRYKFR---QRVEIDKEDHSPVTIADRAEAEAMSVILKSF 131
Db 6 EQLRLIAQAAVSGEILIQYFRSHLQGGTKIDQVS-AIVTQADEAEQAQVHILIQAF 64
```

```
QY 132 PTHAIFGEENGWRCANSAADFVWVLDPIDGTSFTGKPLFGLTIALLLH-NGKPVIGVID 190
Db 65 PODGVIREB-GKNIAGKSG-YTWVLDPIDGTSFSSVGRGLPIFATLGLVDADMRPVLGIAH 122

QY 191 OPIRLRERWIGVDGKQTTLNG----QEISVRSNLLAQAYLYTTSPLHFEADA-EDAFIRV 245
Db 123 QPISGDRWOGVQGEOSNVNGIPLVNPYKASEINLTA-ACIVSTTPLMETTPVQOQKMAI 181

QY 246 RNKVVVPLVGCDCYAYALLASGFVD--IVVESGLKPYDFLSLPVPVIEGAGSGITDWRGD 302
Db 182 YRQCORTAFGDCFRNYLSAASGWTAMPLVIVEADLNFDYFCALIPILTGANYCYFTDMQOK 241

QY 303 KLRWPVTAESRPTSFNVVAAGDARVHKEALDALR 336
Db 242 EL-----TPESTEVSASPNPKLHSELLAFIQ 267

RESULT 10
Q988Q3 PRELIMINARY; PRT; 239 AA.

AC Q988Q3;
DT 01-OCT-2001 (TREMBLrel. 18, Created)
DT 01-OCT-2001 (TREMBLrel. 18, Last sequence update)
DT 01-OCT-2001 (TREMBLrel. 18, Last annotation update)
DE N-anidino-scyllino-samine-4-phosphate phosphatase.
GN MLL6642.
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Phyllobacteriaceae; Mesorhizobium.
OX NCBI_TaxID=381;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=MAFF303099;
RX MEDLINE=21082930; PubMed 11214968;
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA Watanabe A., Idegawa K., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RL DNA Res. 7:331-338(2000).
DR EMBL: AF003009; BAB52894.1; -.
DR InterPro: IPR000760; Inositol_P.
DR Pfam: PF00459; inositol_P; 1.
KW Complete proteome.
SQ SEQUENCE 239 AA; 26168 MW; EBFKA2525C4554763 CRC64;

Query Match 18.1%; Score 417; DB 16; Length 239;
Best Local Similarity 47.4%; Pred. No. 1.1e-17;
Matches 72; Conservative 35; Mismatches 72; Indels 14; Gaps 5;

QY 84 RAADAAGEVLRYKFRQREVEIDKED---HSPVTIADRAEAEAMSVILKSFPTHAIFGE 139
Db 12 RLADAAANVALAYRYKPIEVEEKVKPGYRDPVPTIADKRAEQVIRALISGEFFDHAVLGE 71

QY 140 ENGWRCANSAADFVWVLDPIDGTSFTGKPLFGLTIALLLHNGKPVIGVIDQPIRLR-W 198
Db 72 EFG---ESGGGPVKWIDPIDGTRPFCGIPVWGTILGLTVDGKACMGMSQPTGEREW 128

QY 199 IGVGDKQTTLNG----QEISVRSNLLAQAYLYTTSPLHFEADAEDAFIRVKNKVPLY 254
Db 129 --ATGTEAWTGGASGSRLOTKAAGSLATALLHNSPDREPDFDINFARLKDAAALMTRY 186

QY 255 GDCYAYALIASG 267
Db 187 GGEYAFAMLAAG 199

RESULT 11
Q53743 PRELIMINARY; PRT; 273 AA.

ID Q53743
```



```

QY 78 LVEAQAADAAGEVLRKYF--ORVEIIDKEDHSPVTIADREAEAMVSVILKSPFTHA 135
DB 4 MLTIAIRAARKAGNLIANKVETPDVAEASOKGSDNFVTVNDREHLIDIVIKRSPKHT 63
QY 136 IFGEENGWCAENSADFVWVLDPIDGTSFKITGPKLFGTLIALHNGKPVIGVIDQPILR 195
DB 64 IISEEGELVGEDD-DVQWVIDPLDGTTFIKRLPHFAVSIAVRIKGRTEVAVVYDPMRN 122
QY 196 ERWIGVDGKOTTLNGQEISVRSN-----LLAQYLYTTSPLHFEADAEDAFIRVRNK 248
DB 123 ELFTASRGQAQLNGYRL--RGTNAKDLDTGTLATGPFKVKQH-----APAYIRVWGK 174
QY 249 V-----KVPLGDCYAYALLASGFVDIVVESGLKPYDFLSLVPVIEGAGGSITDWRGK 303
DB 175 LFEQCADFRITGSAALDLAYVAAGRVDGFEIGLKPWFAGGELLVRESGGIVTDAGGH 234
QY 304 LHPVPTAESRPTSFNVVAAGDARVHKEALDAL 335
DB 235 NHF-----SSGNIV-AGNPRIVKSIVQAM 257

RESULT 14
O32889
ID O32889 PRELIMINARY; PRT; 255 AA.
AC O32889;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE MONOPHOSPHATASE (Putative monophosphatase).
GN ML0662 OR MLCB1779.29.
OS Mycobacterium leprae.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae.
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_taxid=1769;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TN;
RX MEDLINE=21128732; PubMed=11234002;
RA Cole S.T., Eiglmeyer K., Parkhill J., James K.D., Thomson N.R.,
RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Barrell B.G.;
RA "Massive gene decay in the leprosy bacillus.";
RL Nature 409:1007-1011(2001).
DR EMBL; Z98271; CAB11010.1; -.
DR EMBL; AL583919; CAC30171.1; -.
DR Leproma; ML0662; -.
DR InterPro; IPR000760; Inositol_P.
DR Pfam; PF00459; inositol_P; 1.
DR PROSITE; PS00629; IMP_1; 1.
KW Complete proteome.
SQ SEQUENCE 255 AA; 27040 MW; 2633DDC72324D545 CRC64;

Query Match 16.4%; Score 283; DB 16; Length 255;
Best Local Similarity 31.4%; Pred. No. 6.5e-15;
Matches 85; Conservative 40; Mismatches 102; Indels 44; Gaps 9;

QY 81 VAQRAADAAGEVLRKYF-RORVEIIDKEDHSPVTIADREAEAMVSVILKSPFTHAIFGE 139
DB 2 LAUTLADRADALTSARFGALNLVRDPLTPVTDRAVEADRAVLGRERPKDGILGE 61

QY 140 ENGRWCAENSADFVWVLDPIDGTSFKITGPKLFGTLIALHNGKPVIGVIDQPILRERW- 198
DB 62 EYGGTITFSQO--WIVDPIDGTGKNEVRGVPVWASLIALLLEDGVPISGVVSAPALQRWW 119

QY 199 -----IGVDG-----KOTTLNGQEISVRSNLLAQYLYTTSPLHFEADAEDAF 242
DB 235 NHF-----SSGNIV-AGNPRIVKSIVQAM 257

```

```

DB 120 AARGQAFVAVDGVPRRLAVSEVADLNSASLSFSSLSGNAQRGL-----RDRE 167
QY 243 IRRNRKV-KVPLGDCYAYALLASGFVDIVVESGLKPYDFLSLVPVIEGAGGSITDWRG 301
DB 168 LELTDVAVRVRAG-DFLSYCLLAEGAIDVAAEPKVSVDLAALDIVVREAGGVLTGLDG 236
QY 302 DXLHWPVTAESRPTSFNVVAAGDARVHKEAL 332
DB 227 -----TPGPHGGSAVAT-NCTHQTIVL 247

RESULT 15
O9K4B1
ID O9K4B1 PRELIMINARY; PRT; 266 AA.
AC O9K4B1;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Putative monophosphatase.
GN SCO5208 OR SC7E4.05C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae.
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_taxid=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Seeger K.J., Harris D.;
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kieser H.M., Denapaite D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RA "A set of ordered contigs and a detailed genetic and physical map for
RL the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neil S.,
RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RA "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2)";
RL Nature 417:141-147(2002).
DR EMBL; AL359214; CAB94593.1; -.
DR HSSP; P29218; IIMP.
DR InterPro; IPR002106; AAIRNA_ligaseII.
DR InterPro; IPR000760; Inositol_P.
DR Pfam; PF00459; inositol_P; 1.
DR PROSITE; PS00339; AA_TRNA_LIGASE_IL_2; UNKNOWN_1.
DR PROSITE; PS00629; IMP_1; 1.
SQ SEQUENCE 266 AA; 29083 MW; 9C2802DB177B243E CRC64;

Query Match 16.1%; Score 278.5; DB 16; Length 266;
Best Local Similarity 30.3%; Pred. No. 1.6e-14;
Matches 83; Conservative 51; Mismatches 111; Indels 29; Gaps 10;

QY 77 RLVEAQAADAAGEVLRKYFRORVEIIDKEDHSPVTIADREAEAMVSVILKSPFTHAI 136
DB 235 NHF-----SSGNIV-AGNPRIVKSIVQAM 257

```

Db 9 RLAHVLADAADA--TMDRFKALDLKVKTPDQMTVPVSEADKAAEELIRGHLSRARPRDSV 66
QY 137 FGEENGWRCAEENSADFWVWLDPIDGTFKSFITGKPLFGTLLIALLH---NGKPVIGVIDOP 192
Db 67 HGEFEG---VAGTGPRRWVIDPIDGTFKNVRCVVPVWATLIALMEAKEGGYQPVVGLVSAP 123
QY 193 ILRERWIGVD-----GKOTTLNGOEISVRSCNLLAQAA-YLYTTSPHLFEADAEDAFIRV 245
Db 124 ALGRRWVAVEDHCAFTGRSLT-SAHLRHVSQVSTLSDASFAYSSLSCWEEQGRLDGFLDL 182
QY 246 RNKV-KVPLYGCDYAYALLASGFVDIVVESGLKPYDFLSLVPVIEGAGGSITDWRGDKL 304
Db 183 TREVMRTRAYG-DFWPMVMAEGSVDLCAEPELSMDMAANAIIIVTEAGGTFTGLDG--- 238
QY 305 HWPVTAESRPTSFNVVAAGDARVHKALDALRWR 338
Db 239 -----RPGPHSGN-AAASNGRLHDELLGYLNQR 265

Search completed: November 17, 2002, 16:41:29
b time : 95 secs

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 30, 2002, 17:13:04 ; Search time 28.74 Seconds
(without alignments)
2034.526 Million cell updates/sec

Title: US-09-686-522a-14
Perfect score: 1730
Sequence: 1 MLSSSSSTHSDFSPFGLAS.....VVAAGDARVHKEALDRLWR 338

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

otal number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1020	59.0	286	10 Q9T021	Q9T021 arabidopsis
2	551	31.8	259	16 Q92M31	Q92M31 rhizobium m
3	483	27.9	308	16 Q930E4	Q930E4 rhizobium m
4	477	27.6	278	16 Q9A643	Q9A643 caulobacter
5	460	26.6	263	2 Q9KWB5	Q9KWB5 agrobacteri
6	449	26.0	262	16 Q98H43	Q98H43 rhizobium l
7	396.5	22.9	267	16 P73806	P73806 synecocyst
8	317	18.3	239	16 Q988Q3	Q988Q3 rhizobium l
9	317	18.3	273	2 Q53743	Q53743 streptomyce
10	282	16.9	271	16 Q9HX14	Q9HX14 pseudomonas
11	287	16.6	267	16 Q9CNV8	Q9CNV8 pasteurilla
12	283	16.4	255	16 Q32889	Q32889 mycobacteri
13	279	16.1	264	16 Q67791	Q67791 aquilex aeo
14	278.5	16.1	266	2 Q9K4B1	Q9K4B1 streptomyce
15	273.5	15.8	260	16 P95189	P95189 mycobacteri
16	265.5	15.3	271	10 Q9M8S8	Q9M8S8 arabidopsis

17	258.5	14.9	371	10 Q94F00	Q94F00 arabidopsis
18	248	14.3	288	16 Q9KTY5	Q9KTY5 vibrio chol
19	239.5	13.8	286	17 Q9HQ12	Q9HQ12 halobacteri
20	237	13.7	272	16 Q92RQ1	Q92RQ1 rhizobium m
21	234	13.5	275	16 Q9PAM0	Q9PAM0 xyliella fas
22	231.5	13.4	353	10 Q9SA15	Q9SA15 arabidopsis
23	230	13.3	221	2 P71487	P71487 methyloicr
24	230	13.3	287	16 P74542	P74542 synecocyst
25	229	13.2	284	16 Q98D39	Q98D39 rhizobium l
26	228	13.2	286	16 Q98F59	Q98F59 rhizobium l
27	221	12.8	266	16 Q92WJ1	Q92WJ1 rhizobium m
28	221	12.8	272	2 Q30546	Q30546 agrobacteri
29	220	12.7	275	16 Q98NP3	Q98NP3 rhizobium l
30	220	12.7	294	16 Q92X03	Q92X03 rhizobium m
31	219	12.7	290	11 Q91U25	Q91U25 mus musculu
32	218	12.6	319	5 Q23493	Q23493 caenorhabdi
33	215	12.4	255	2 Q9R6D4	Q9R6D4 agrobacteri
34	213	12.3	264	2 Q9S3X5	Q9S3X5 streptomyce
35	212.5	12.3	277	11 Q9D066	Q9D066 mus musculu
36	212	12.3	284	5 Q9VUW2	Q9VUW2 drosophila
37	211	12.2	266	16 Q92M71	Q92M71 rhizobium m
38	210.5	12.2	288	5 Q9VUW4	Q9VUW4 drosophila
39	209	12.1	256	2 P72460	P72460 streptomyce
40	209	12.1	257	16 Q92CW7	Q92CW7 listeria in
41	208.5	12.1	277	11 Q924B0	Q924B0 mus musculu
42	207.5	12.0	127	2 Q69216	Q69216 azotobacter
43	207	12.0	254	17 Q9V1Z6	Q9V1Z6 pyrococcus
44	207	12.0	397	10 Q9M0Y6	Q9M0Y6 arabidopsis
45	206.5	11.9	287	2 Q52855	Q52855 rhizobium l

ALIGNMENTS

RESULT 1

Q9T021 ID Q9T021 PRELIMINARY; PRT; 286 AA.
AC Q9T021;
DT 01-MAY-2000 (TREMblrel, 13, Created)
DT 01-MAY-2000 (TREMblrel, 13, Last sequence update)
DT 01-JUN-2001 (TREMblrel, 17, Last annotation update)
DE INOSITOL MONOPHOSPHATASE-LIKE PROTEIN.
GN T22F8.20 OR A14G39120.
OS Arabidopsis thaliana (Mouse-ear cross).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Bevan M., Zimmermann W., Gruenisen A., Wambutt R., Bancroft I.,
RA Meves H.W., Mayer K.F.X., Schueller C.;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Zimmermann W., Gruenisen A., Wambutt R., Kalicki J., Wohldmann P.,
RA Smith A., Meves H.W., Lemcke K., Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL050351; CAB43627.1; .
DR EMBL: AL161594; CAB80575.1; .
DR InterPro: IPR000760; Inositol_P.
DR Pfam: PF00459; Inositol_P; 1.
DR PROSITE: PS00629; IMP_1; 1.
SQ SEQUENCE 286 AA; 31635 MW; 5C2BB2BA90FD10BB CRC64;

DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
 DE INOSITOL MONOPHOSPHATASE FAMILY PROTEIN.
 GN CC2252.
 OS Caulobacter crescentus.
 OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
 OC Caulobacter.
 OX NCBI_TaxID=69394;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 19089 / CB15;
 RX MEDLINE=21173698; PubMed=11259647;
 RA Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
 RA Eisen J., Heidelberg J.F., Alvey M.R.K., Ohta N., Maddock J.R.,
 RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
 RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
 RA Kolony J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
 RA Uterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
 RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
 IT "Complete genome sequence of Caulobacter crescentus.";
 J. Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
 DR EMBL; AE005895; AAK24223.1; -;
 DR TIGR; CC2252; -;
 DR InterPro: IPR000760; Inositol_P.
 DR Pfam: PF00459; inositol_P; 2.
 DR Complete proteome.
 KW SEQUENCE 278 AA; 29688 MW; 281299D0F028757A CRC64;
 SQ
 Query Match 27.6%; Score 477; DB 16; Length 278;
 Best Local Similarity 40.8%; Pred. No. 9.6e-31;
 Matches 116; Conservative 48; Mismatches 88; Indels 32; Gaps 9;
 QY 73 LATERLVEAQRAD---AAGEVLRYKFORVEIIDK-----EDHS---PVTIADRE 118
 Db 3 LSADRLAALDAFTLDLRASADVILFLFRADHGLDKG:GKNLPDTHAAFDVPTADRG 62
 QY 119 AEEAMSVILKSPPTTAIFGEENGWRCAN--SADFVWVLDPIIDGTSFKITGKPLFCTLI 176
 Db 63 AEAIRALIAQRPDHCVIGEEVG---EDRPAEFVWVLDPIIDGTRAFIAGLPLWTTLI 118
 QY 177 ALLHNGKPVITGVIDQPLIRERWIGVDGKQITLNGQE---ISVRSNLLAQAYLYTTSPLH 232
 Db 119 GLRHEGRVPLGSGQPVYNEIFIGHAGGARLVSGGEARPIRVRECANINDAVIATDPDA 178
 QY 233 LFEADAEDAFIRVNRKVKVPLGYCDCYAYALLASGFVDIVVESGLKPYDPLSLVVPVIEGA 292
 Db 179 CFGAERGAWLQVRAAKLARLGCDAAYANVAMGRMDWVIEAGLKSWDIEAAILPILEGA 238
 JY 293 GGSITDWRGDKLHPVTAESRPTSFNVVAAGDARVHKEALDALR 336
 Db 239 GGMVTNWRGE---PV---GPNQGQWVIGDRRPLDEALVSLR 274
 RESULT 5
 Q9KWB5 PRELIMINARY; PRT; 263 AA.
 AC Q9KWB5;
 DT 01-OCT-2000 (Tremblrel. 15, Created)
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
 DE RIORF69 PROTEIN.
 GN RIORF69.
 OS Agrobacterium rhizogenes.
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Rhizobiaceae; Rhizobium.
 OX NCBI_TaxID=359;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-MAFF03-01724; PLASMID=PRI1724;
 RA Moriguchi K., Maeda Y., Satou M., Kataoka M., Tanaka N., Yoshida K.;
 IT "Analysis of unique variable region of a plant root inducing plasmid,"

RT pr11724, by the construction of its physical map and library.";
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-MAFF03-01724; PLASMID=PRI1724;
 RA Moriguchi K., Maeda Y., Satou M., Kataoka M., Tanaka N., Yoshida K.;
 IT "The complete nucleotide sequence of a Ri (root inducing) plasmid
 RT indicates its chimerical structure between Ti and Sym plasmids.";
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN-MAFF03-01724; PLASMID=PRI1724;
 RA Moriguchi K., Maeda Y., Satou M., Kataoka M., Tanaka N., Yoshida K.;
 IT "Analysis of unique variable region of a plant root inducing plasmid,
 RT pr11724, by the construction of its physical map and library.";
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN-MAFF03-01724; PLASMID=PRI1724;
 RX MEDLINE=20241294; PubMed=10780382;
 RA Maeda Y., Moriguchi K., Kataoka M., Satou M., Satuti N., Tanaka N.,
 RA Yoshida K.;
 IT "Genome structure of Ri plasmid (1): Sequencing analysis of T-DNA and
 RT its flanking regions of pr11724 in Japanese Agrobacterium
 RT rhizogenes.";
 RL Nucleic Acids Symp. Ser. 42:67-68(1999).
 DR EMBL; AB039932; BAA97780.1; -;
 DR EMBL; AP002086; BAB16188.1; -;
 DR HSSP; P29218; 11MF.
 DR InterPro: IPR000760; Inositol_P.
 DR Pfam: PF00459; inositol_P; 1.
 DR PROSITE; PS00629; IMP_1; 1.
 KW Plasmid.
 SQ SEQUENCE 263 AA; 28110 MW; 2979EA9F8958A25F CRC64;
 Query Match 26.6%; Score 460; DB 2; Length 263;
 Best Local Similarity 38.8%; Pred. No. 2.1e-29;
 Matches 107; Conservative 55; Mismatches 82; Indels 32; Gaps 8;
 QY 74 ATERLVE---VAGRAAAGAEVLKRYFORVEIIDK--EDHSPTTIADREAEAMYSVIL 128
 Db 3 STEHWISPDPIRKIAAFAAETLPR-FRLGIVVVKIAGCFDPTVTEADRETEKALALIR 61
 QY 129 KSPPTTAIFGEENGWRCAN--SADFVWVLDPIIDGTSFKITGKPLFCTLIALLHNGKPVIGV 188
 Db 62 SEYPGHGVLGEEYGNAGD--AELVWVIDPIDCTRSFISGIPLWGLTGLTRDGDVAGM 119
 QY 189 IDQPIILRERWIGVDGKQITLNGQE---QEISVRSNLLAQAYLYTTSPLHFEADRE 239
 Db 120 MAQPFITGELF-----YSTGNGAFYEGPHQQTQATSKDLADATLFTTTPAIFADRR 173
 QY 240 DAFIRVNRKVKVPLGYCDCYAYALLASGFVDIVVESGLKPYDPLSLVVPVIEGAGSITDW 299
 Db 174 SSYDLLEKQVRLARYGTDYAYCYMLAAGIDCVVEAGLYPYDIALIPIIEQAGGVVTDW 233
 QY 300 RGDKLHPVTAESRPTSFNVVAAGDARVHKEALDAL 335
 Db 234 DGG-----PAE---RGGGIVAAANATLHAQIMACL 260
 RESULT 6
 Q98H43 PRELIMINARY; PRT; 262 AA.
 ID Q98H43
 AC Q98H43;

DT	01-OCT-2001 (TReMBLrel. 18, Created)	
DT	01-OCT-2001 (TReMBLrel. 18, Last sequence update)	
DT	01-OCT-2001 (TReMBLrel. 18, Last annotation update)	
DN	MONO-PHOSPHATASE.	
GN	MLL3039.	
OS	Rhizobium loti (Mesorhizobium loti).	
OC	Bacteria, Proteobacteria, alpha subdivision; Rhizobiaceae group;	
OC	Phyllobacteriaceae; Mesorhizobium.	
ON	NCBI_TaxID=381;	
OX	{1}	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=MAFF303099;	
RX	MEDLINE=21082930; PubMed=11214968;	
RA	Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,	
RA	Watanabe A., Iidesawa K., Ishikawa A., Kawashima K., Kimura T.,	
RA	Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,	
RA	Mochizuki Y., Nakayama S., Nakazaki N., Shimpō S., Sugimoto M.,	
RA	Takeuchi C., Yamada M., Tabata S.;	
RT	"Complete genome structure of the nitrogen-fixing symbiotic bacterium	
RT	Mesorhizobium loti.";	
TA	DNA Res. 7:331-338(2000).	
AC	EMBL; AP003001; BAB50023.1; -	
DR	InterPro: IPR000760; Inositol_P.	
DR	Pfam: PF00459; inositol_P; 1.	
KW	Complete proteome.	
SQ	SEQUENCE 262 AA; 27892 MW; E38C73FB36D27CE5 CRC64;	
Query Match 26.0%; Score 449; DB 16; Length 262;		
Best Local Similarity 39.8%; Pred. No. 1.7e-28;		
Matches 105; Conservative 50; Mismatches 91; Indels 18; Gaps		
Qy	79 VEVAQRAADAAGVLRYKRYQRVEIIDKEDHS--PVTIADREAEAMVSVILKSPPTTHAI 136	
Db	::: ::::: : : :	
Qy	10 IDFMRIIAQAAAAETLPRFAAGAVANKGSGPDVTEADRETERAIRALIAAQYPDHGI 69	
Qy	137 FGEENGHRCAEEN-SADQVWVLDPIDGCTKSFITCKPLFGTLIALLHNGKPVIGVIDOPILR 195	
Db	: : : : : : : : : : : : : : : :	
Qy	70 LGEEHG--SENTSSRHVVWIDPIDCTRAFISGLPVWGI.VGLTVDGDVAGMMAQPTFG 126	
Qy	196 ER-WIGVDGKQTTLNG--QEISVRSNCLLAQAYLYTTSPLHFEADAEDAFIRVRNKVKVP 252	
Db	::: ::::: : : :	
Qy	127 ELFVANASGSHYEGPGPKLSTKTRTRDDEATLFTTTFALFKGEARTYDAFEKQVQLA 186	
Qy	253 LYGCDCYAYALLASGFVDIVVESGLKPYDFLSVPVIEGAGSGITDWCGLKHWVPVTAES 312	
Db		
Qy	187 RYGDCAFAMIASGSDIVADPLGPKYDIVALPIIERAGGVVITFDGG-----PAEK 240	
Qy	313 RPTSFNVVAAGDARVHKALDALR 336	
Qy	::: : : : : : : : : : : : : : : :	
Qy	241 --GGDVLAATAATPELHAAANAAALR 261	
Qy	PRELIMINARY; PRT; 267 AA.	
RESULT	7	
ID	P73806	
AC	P73806;	
DT	01-FEB-1997 (TReMBLrel. 02, Created)	
DT	01-FEB-1997 (TReMBLrel. 02, Last sequence update)	
DT	01-DEC-2001 (TReMBLrel. 19, Last annotation update)	
DE	EXTRAGENIC SUPPRESSOR.	
GN	SUHH OR SLL1959.	
OS	Synechocystis sp. (strain PCC 6803).	
OC	Bacteria; Cyanobacteria; Chroococcales; Synechocystis.	
OX	NCBI_TaxID=1148;	
RN	{1}	
RP	SEQUENCE FROM N.A.	
RX	MEDLINE=97061201; PubMed=8905231;	
RA	Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,	
RA	Miyajima N., Hirosewa M., Sugiyama M., Sasamoto S., Kimura T.,	
RA	Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,	
RA	Shimpō S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,	
RA	Tabata S.;	

```
QY 84 RAADAAGEVLRYFORVEIIDKED----HSPVTIADREAEAMVSVILKSPFTHAIFCE 139
+||||| |:|: |:|: |||||::|:: |:|: |||||:|
Db 12 RLADAAANVALAYRKPIVEKEKVPGRFDPVTIADKAEQVIRALISGEFPDHAVLGE 71
+||||| |:|: |:|: |||||::|:: |:|: |||||:|
QY 140 ENGWRCAENSADFWVWLDPIDGKTSGFITGKPLFGTLIALHNGKVPVIGVIDQPILRR-W 198
+||||| |:|: |:|: |||||::|:: |:|: |||||:|
Db 72 EFG---ESGQGPVKWIIDPIDGTCIGIPVMGTGLIGLTVGDKACMGMSQPFGTGERFW 128
+||||| |:|: |:|: |||||::|:: |:|: |||||:|
QY 199 IGVDCQKQTTLNG---QEISVRSNLLAQAYLYTTSPLHFEADAEDAFIRVNKKVPLY 254
+||||| |:|: |:|: |||||::|:: |:|: |||||:|
Db 129 --ATGTEAWTGAGSSRLQTKAAGSLATAIUHTNSPDRFPDPFDINFARLKDAALMTRY 186
+||||| |:|: |:|: |||||::|:: |:|: |||||:|
QY 255 GDCYAYALLASG 267
+||||| |:|: |:|: |||||::|:: |:|: |||||:|
Db 187 GGEYAFAMLAAG 199
+||||| |:|: |:|: |||||::|:: |:|: |||||:|

RESULT 9
Q53743 PRELIMINARY; PRT; 273 AA.
1 Q53743
2 Q53743
3 Q53743
4 Q53743
5 Q53743
6 Q53743
7 Q53743
8 Q53743
9 Q53743
10 Q53743
11 Q53743
12 Q53743
13 Q53743
14 Q53743
15 Q53743
16 Q53743
17 Q53743
18 Q53743
19 Q53743
20 Q53743
21 Q53743
22 Q53743
23 Q53743
24 Q53743
25 Q53743
26 Q53743
27 Q53743
28 Q53743
29 Q53743
30 Q53743
31 Q53743
32 Q53743
33 Q53743
34 Q53743
35 Q53743
36 Q53743
37 Q53743
38 Q53743
39 Q53743
40 Q53743
41 Q53743
42 Q53743
43 Q53743
44 Q53743
45 Q53743
46 Q53743
47 Q53743
48 Q53743
49 Q53743
50 Q53743
51 Q53743
52 Q53743
53 Q53743
54 Q53743
55 Q53743
56 Q53743
57 Q53743
58 Q53743
59 Q53743
60 Q53743
61 Q53743
62 Q53743
63 Q53743
64 Q53743
65 Q53743
66 Q53743
67 Q53743
68 Q53743
69 Q53743
70 Q53743
71 Q53743
72 Q53743
73 Q53743
74 Q53743
75 Q53743
76 Q53743
77 Q53743
78 Q53743
79 Q53743
80 Q53743
81 Q53743
82 Q53743
83 Q53743
84 Q53743
85 Q53743
86 Q53743
87 Q53743
88 Q53743
89 Q53743
90 Q53743
91 Q53743
92 Q53743
93 Q53743
94 Q53743
95 Q53743
96 Q53743
97 Q53743
98 Q53743
99 Q53743
100 Q53743
101 Q53743
102 Q53743
103 Q53743
104 Q53743
105 Q53743
106 Q53743
107 Q53743
108 Q53743
109 Q53743
110 Q53743
111 Q53743
112 Q53743
113 Q53743
114 Q53743
115 Q53743
116 Q53743
117 Q53743
118 Q53743
119 Q53743
120 Q53743
121 Q53743
122 Q53743
123 Q53743
124 Q53743
125 Q53743
126 Q53743
127 Q53743
128 Q53743
129 Q53743
130 Q53743
131 Q53743
132 Q53743
133 Q53743
134 Q53743
135 Q53743
136 Q53743
137 Q53743
138 Q53743
139 Q53743
140 Q53743
141 Q53743
142 Q53743
143 Q53743
144 Q53743
145 Q53743
146 Q53743
147 Q53743
148 Q53743
149 Q53743
150 Q53743
151 Q53743
152 Q53743
153 Q53743
154 Q53743
155 Q53743
156 Q53743
157 Q53743
158 Q53743
159 Q53743
160 Q53743
161 Q53743
162 Q53743
163 Q53743
164 Q53743
165 Q53743
166 Q53743
167 Q53743
168 Q53743
169 Q53743
170 Q53743
171 Q53743
172 Q53743
173 Q53743
174 Q53743
175 Q53743
176 Q53743
177 Q53743
178 Q53743
179 Q53743
180 Q53743
181 Q53743
182 Q53743
183 Q53743
184 Q53743
185 Q53743
186 Q53743
187 Q53743
188 Q53743
189 Q53743
190 Q53743
191 Q53743
192 Q53743
193 Q53743
194 Q53743
195 Q53743
196 Q53743
197 Q53743
198 Q53743
199 Q53743
200 Q53743
201 Q53743
202 Q53743
203 Q53743
204 Q53743
205 Q53743
206 Q53743
207 Q53743
208 Q53743
209 Q53743
210 Q53743
211 Q53743
212 Q53743
213 Q53743
214 Q53743
215 Q53743
216 Q53743
217 Q53743
218 Q53743
219 Q53743
220 Q53743
221 Q53743
222 Q53743
223 Q53743
224 Q53743
225 Q53743
226 Q53743
227 Q53743
228 Q53743
229 Q53743
230 Q53743
231 Q53743
232 Q53743
233 Q53743
234 Q53743
235 Q53743
236 Q53743
237 Q53743
238 Q53743
239 Q53743
240 Q53743
241 Q53743
242 Q53743
243 Q53743
244 Q53743
245 Q53743
246 Q53743
247 Q53743
248 Q53743
249 Q53743
250 Q53743
251 Q53743
252 Q53743
253 Q53743
254 Q53743
255 Q53743
256 Q53743
257 Q53743
258 Q53743
259 Q53743
260 Q53743
261 Q53743
262 Q53743
263 Q53743
264 Q53743
265 Q53743
266 Q53743
267 Q53743
268 Q53743
269 Q53743
270 Q53743
271 Q53743
272 Q53743
273 Q53743
274 Q53743
275 Q53743
276 Q53743
277 Q53743
278 Q53743
279 Q53743
280 Q53743
281 Q53743
282 Q53743
283 Q53743
284 Q53743
285 Q53743
286 Q53743
287 Q53743
288 Q53743
289 Q53743
290 Q53743
291 Q53743
292 Q53743
293 Q53743
294 Q53743
295 Q53743
296 Q53743
297 Q53743
298 Q53743
299 Q53743
300 Q53743
301 Q53743
302 Q53743
303 Q53743
304 Q53743
305 Q53743
306 Q53743
307 Q53743
308 Q53743
309 Q53743
310 Q53743
311 Q53743
312 Q53743
313 Q53743
314 Q53743
315 Q53743
316 Q53743
317 Q53743
318 Q53743
319 Q53743
320 Q53743
321 Q53743
322 Q53743
323 Q53743
324 Q53743
325 Q53743
326 Q53743
327 Q53743
328 Q53743
329 Q53743
330 Q53743
331 Q53743
332 Q53743
333 Q53743
334 Q53743
335 Q53743
336 Q53743
337 Q53743
338 Q53743
339 Q53743
340 Q53743
341 Q53743
342 Q53743
343 Q53743
344 Q53743
345 Q53743
346 Q53743
347 Q53743
348 Q53743
349 Q53743
350 Q53743
351 Q53743
352 Q53743
353 Q53743
354 Q53743
355 Q53743
356 Q53743
357 Q53743
358 Q53743
359 Q53743
360 Q53743
361 Q53743
362 Q53743
363 Q53743
364 Q53743
365 Q53743
366 Q53743
367 Q53743
368 Q53743
369 Q53743
370 Q53743
371 Q53743
372 Q53743
373 Q53743
374 Q53743
375 Q53743
376 Q53743
377 Q53743
378 Q53743
379 Q53743
380 Q53743
381 Q53743
382 Q53743
383 Q53743
384 Q53743
385 Q53743
386 Q53743
387 Q53743
388 Q53743
389 Q53743
390 Q53743
391 Q53743
392 Q53743
393 Q53743
394 Q53743
395 Q53743
396 Q53743
397 Q53743
398 Q53743
399 Q53743
400 Q53743
401 Q53743
402 Q53743
403 Q53743
404 Q53743
405 Q53743
406 Q53743
407 Q53743
408 Q53743
409 Q53743
410 Q53743
411 Q53743
412 Q53743
413 Q53743
414 Q53743
415 Q53743
416 Q53743
417 Q53743
418 Q53743
419 Q53743
420 Q53743
421 Q53743
422 Q53743
423 Q53743
424 Q53743
425 Q53743
426 Q53743
427 Q53743
428 Q53743
429 Q53743
430 Q53743
431 Q53743
432 Q53743
433 Q53743
434 Q53743
435 Q53743
436 Q53743
437 Q53743
438 Q53743
439 Q53743
440 Q53743
441 Q53743
442 Q53743
443 Q53743
444 Q53743
445 Q53743
446 Q53743
447 Q53743
448 Q53743
449 Q53743
450 Q53743
451 Q53743
452 Q53743
453 Q53743
454 Q53743
455 Q53743
456 Q53743
457 Q53743
458 Q53743
459 Q53743
460 Q53743
461 Q53743
462 Q53743
463 Q53743
464 Q53743
465 Q53743
466 Q53743
467 Q53743
468 Q53743
469 Q53743
470 Q53743
471 Q53743
472 Q53743
473 Q53743
474 Q53743
475 Q53743
476 Q53743
477 Q53743
478 Q53743
479 Q53743
480 Q53743
481 Q53743
482 Q53743
483 Q53743
484 Q53743
485 Q53743
486 Q53743
487 Q53743
488 Q53743
489 Q53743
490 Q53743
491 Q53743
492 Q53743
493 Q53743
494 Q53743
495 Q53743
496 Q53743
497 Q53743
498 Q53743
499 Q53743
500 Q53743
501 Q53743
502 Q53743
503 Q53743
504 Q53743
505 Q53743
506 Q53743
507 Q53743
508 Q53743
509 Q53743
510 Q53743
511 Q53743
512 Q53743
513 Q53743
514 Q53743
515 Q53743
516 Q53743
517 Q53743
518 Q53743
519 Q53743
520 Q53743
521 Q53743
522 Q53743
523 Q53743
524 Q53743
525 Q53743
526 Q53743
527 Q53743
528 Q53743
529 Q53743
530 Q53743
531 Q53743
532 Q53743
533 Q53743
534 Q53743
535 Q53743
536 Q53743
537 Q53743
538 Q53743
539 Q53743
540 Q53743
541 Q53743
542 Q53743
543 Q53743
544 Q53743
545 Q53743
546 Q53743
547 Q53743
548 Q53743
549 Q53743
550 Q53743
551 Q53743
552 Q53743
553 Q53743
554 Q53743
555 Q53743
556 Q53743
557 Q53743
558 Q53743
559 Q53743
560 Q53743
561 Q53743
562 Q53743
563 Q53743
564 Q53743
565 Q53743
566 Q53743
567 Q53743
568 Q53743
569 Q53743
570 Q53743
571 Q53743
572 Q53743
573 Q53743
574 Q53743
575 Q53743
576 Q53743
577 Q53743
578 Q53743
579 Q53743
580 Q53743
581 Q53743
582 Q53743
583 Q53743
584 Q53743
585 Q53743
586 Q53743
587 Q53743
588 Q53743
589 Q53743
590 Q53743
591 Q53743
592 Q53743
593 Q53743
594 Q53743
595 Q53743
596 Q53743
597 Q53743
598 Q53743
599 Q53743
600 Q53743
601 Q53743
602 Q53743
603 Q53743
604 Q53743
605 Q53743
606 Q53743
607 Q53743
608 Q53743
609 Q53743
610 Q53743
611 Q53743
612 Q53743
613 Q53743
614 Q53743
615 Q53743
616 Q53743
617 Q53743
618 Q53743
619 Q53743
620 Q53743
621 Q53743
622 Q53743
623 Q53743
624 Q53743
625 Q53743
626 Q53743
627 Q53743
628 Q53743
629 Q53743
630 Q53743
631 Q53743
632 Q53743
633 Q53743
634 Q53743
635 Q53743
636 Q53743
637 Q53743
638 Q53743
639 Q53743
640 Q53743
641 Q53743
642 Q53743
643 Q53743
644 Q53743
645 Q53743
646 Q53743
647 Q53743
648 Q53743
649 Q53743
650 Q53743
651 Q53743
652 Q53743
653 Q53743
654 Q53743
655 Q53743
656 Q53743
657 Q53743
658 Q53743
659 Q53743
660 Q53743
661 Q53743
662 Q53743
663 Q53743
664 Q53743
665 Q53743
666 Q53743
667 Q53743
668 Q53743
669 Q53743
670 Q53743
671 Q53743
672 Q53743
673 Q53743
674 Q53743
675 Q53743
676 Q53743
677 Q53743
678 Q53743
679 Q53743
680 Q53743
681 Q53743
682 Q53743
683 Q53743
684 Q53743
685 Q53743
686 Q53743
687 Q53743
688 Q53743
689 Q53743
690 Q53743
691 Q53743
692 Q53743
693 Q53743
694 Q53743
695 Q53743
696 Q53743
697 Q53743
698 Q53743
699 Q53743
700 Q53743
701 Q53743
702 Q53743
703 Q53743
704 Q53743
705 Q53743
706 Q53743
707 Q53743
708 Q53743
709 Q53743
710 Q53743
711 Q53743
712 Q53743
713 Q53743
714 Q53743
715 Q53743
716 Q53743
717 Q53743
718 Q53743
719 Q53743
720 Q53743
721 Q53743
722 Q53743
723 Q53743
724 Q53743
725 Q53743
726 Q53743
727 Q53743
728 Q53743
729 Q53743
730 Q53743
731 Q53743
732 Q53743
733 Q53743
734 Q53743
735 Q53743
736 Q53743
737 Q53743
738 Q53743
739 Q53743
740 Q53743
741 Q53743
742 Q53743
743 Q53743
744 Q53743
745 Q53743
746 Q53743
747 Q53743
748 Q53743
749 Q53743
750 Q53743
751 Q53743
752 Q53743
753 Q53743
754 Q53743
755 Q53743
756 Q53743
757 Q53743
758 Q53743
759 Q53743
760 Q53743
761 Q53743
762 Q53743
763 Q53743
764 Q53743
765 Q53743
766 Q53743
767 Q53743
768 Q53743
769 Q53743
770 Q53743
771 Q53743
772 Q53743
773 Q53743
774 Q53743
775 Q53743
776 Q53743
777 Q53743
778 Q53743
779 Q53743
780 Q53743
781 Q53743
782 Q53743
783 Q53743
784 Q53743
785 Q53743
786 Q53743
787 Q53743
788 Q53743
789 Q53743
790 Q53743
791 Q53743
792 Q53743
793 Q53743
794 Q53743
795 Q53743
796 Q53743
797 Q53743
798 Q53743
799 Q53743
800 Q53743
801 Q53743
802 Q53743
803 Q53743
804 Q53743
805 Q53743
806 Q53743
807 Q53743
808 Q53743
809 Q53743
810 Q53743
811 Q53743
812 Q53743
813 Q53743
814 Q53743
815 Q53743
816 Q53743
817 Q53743
818 Q53743
819 Q53743
820 Q53743
821 Q53743
822 Q53743
823 Q53743
824 Q53743
825 Q53743
826 Q53743
827 Q53743
828 Q53743
829 Q53743
830 Q53743
831 Q53743
832 Q53743
833 Q53743
834 Q53743
835 Q53743
836 Q53743
837 Q53743
838 Q53743
839 Q53743
840 Q53743
841 Q53743
842 Q53743
843 Q53743
844 Q53743
845 Q53743
846 Q53743
847 Q53743
848 Q53743
849 Q53743
850 Q53743
851 Q53743
852 Q53743
853 Q53743
854 Q53743
855 Q53743
856 Q53743
857 Q53743
858 Q53743
859 Q53743
860 Q53743
861 Q53743
862 Q53743
863 Q53743
864 Q53743
865 Q53743
866 Q53743
867 Q53743
868 Q53743
869 Q53743
870 Q53743
871 Q53743
872 Q53743
873 Q53743
874 Q53743
875 Q53743
876 Q53743
877 Q53743
878 Q53743
879 Q53743
880 Q53743
881 Q53743
882 Q53743
883 Q53743
884 Q53743
885 Q53743
886 Q53743
887 Q53743
888 Q53743
889 Q53743
890 Q53743
891 Q53743
892 Q53743
893 Q53743
894 Q53743
895 Q53743
896 Q53743
897 Q53743
898 Q53743
899 Q53743
900 Q53743
901 Q53743
902 Q53743
903 Q53743
904 Q53743
905 Q53743
906 Q53743
907 Q53743
908 Q53743
909 Q53743
910 Q53743
911 Q53743
912 Q53743
913 Q53743
914 Q53743
915 Q53743
916 Q53743
917 Q53743
918 Q53743
919 Q53743
920 Q53743
921 Q53743
922 Q53743
923 Q53743
924 Q53743
925 Q53743
926 Q53743
927 Q53743
928 Q53743
929 Q53743
930 Q53743
931 Q53743
932 Q53743
933 Q53743
934 Q53743
935 Q53743
936 Q53743
937 Q53743
938 Q53743
939 Q53743
940 Q53743
941 Q53743
942 Q53743
943 Q53743
944 Q53743
945 Q53743
946 Q53743
947 Q53743
948 Q53743
949 Q53743
950 Q53743
951 Q53743
952 Q53743
953 Q53743
954 Q53743
955 Q53743
956 Q53743
957 Q53743
958 Q53743
959 Q53743
960 Q53743
961 Q53743
962 Q53743
963 Q53743
964 Q53743
965 Q53743
966 Q53743
967 Q53743
968 Q53743
969 Q53743
970 Q53743
971 Q53743
972 Q53743
973 Q53743
974 Q53743
975 Q53743
976 Q53743
977 Q53743
978 Q53743
979 Q53743
980 Q53743
981 Q53743
982 Q53743
983 Q53743
984 Q53743
985 Q53743
986 Q53743
987 Q53743
988 Q53743
989 Q53743
990 Q53743
991 Q53743
992 Q53743
993 Q53743
994 Q53743
995 Q53743
996 Q53743
997 Q53743
998 Q53743
999 Q53743
1000 Q53743

QY 79 VEVAQRAADAAGEVLRYFORVEIIDKEDHSPVTIADREAEAMVSVILKSPFTHAIF 137
+||||| |:|: |:|: |||||::|:: |:|: |||||:|
Db 7 LELAMRLADATRTTRRFQARDGCGYARKPDRPTVTDATTVEAAVRKESVRAARDDDEFA 66
+||||| |:|: |:|: |||||::|:: |:|: |||||:|
QY 138 GEENGWRCAENSADFWVWLDPIDGKTSGFITGKPLFGTLIALHNGKVPVIGVIDQPILRR 197
+||||| |:|: |:|: |||||::|:: |:|: |||||:|
Db 67 GEETG---GEVTRAGTWIYDPIIDGKTNFLRGYPVWVWATLIALLEDGRPTVGVVAAPALRSR 123
+||||| |:|: |:|: |||||::|:: |:|: |||||:|
QY 198 WTGVDCQKQTTN----GQE---ISVRSNLLAQAYLYTTSPLHFEA-DAEDAFIRVNKKV 249
+||||| |:|: |:|: |||||::|:: |:|: |||||:|
Db 124 WMAAGHGAWLRRGAGAPLRLHVSQVRLNAYLSTNTRTWDAFIRNAYLRLADAC 183
+||||| |:|: |:|: |||||::|:: |:|: |||||:|
QY 250 KVPYGCDCYAYALLASGPDVIVESGLKPYDFLSLVPVIECAGGSITDWRGDKLHWPT 309
+||||| |:|: |:|: |||||::|:: |:|: |||||:|
Db 184 WEDRAFQDFLQCHMVAEGTVDIAAPVVSFMDIAALQILVEEAGGVCTDGLG----- 235
+||||| |:|: |:|: |||||::|:: |:|: |||||:|
QY 310 AESRPTSFNVVAAGDARVHKEALDAL 335
+||||| |:|: |:|: |||||::|:: |:|: |||||:|
Db 236 --GSPRGTCGALSANPELHRLAVAL 259
+||||| |:|: |:|: |||||::|:: |:|: |||||:|

RESULT 10
```

```
Q9HXI4 PRELIMINARY; PRT; 271 AA.
AC Q9HXI4:
DT 01-MAR-2001 (TRENBLrel. 16, Created)
DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE EXTRAGENIC SUPPRESSOR PROTEIN SUHB.
GN PA3818
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID-287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 15642 / PA01;
RX MEDLINE-20437337; PubMed 10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Brinkman F.S.L., Huftaile W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltry L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sater M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen."
RL Nature 406:959-964(2000).
DR EMBL: AE004799; AAC07205.1; -
DR HSSP; P29218; 1MP.
DR InterPro; IPR000760; Inositol_P.
DR Pfam; PF00459; inositol_P; 1.
DR PROSITE; PS00629; IMP_1; 1.
DR PROSITE; PS00630; IMP_2; 1.
KW Complete proteome.
SQ SEQUENCE 271 AA; 29521 MW; EEID4KCFEB871EA CRC64;

Query Match 16.9%; Score 292; DB 16; Length 271;
Best Local Similarity 29.7%; Pred. No. 9.4e-16;
Matches 82; Conservative 54; Mismatches 106; Indels 34; Gaps 8;

QY 78 LVEVAQRAADAAGEVLRYFORVEIIDKEDHSPVTIADREAEAMVSVILKSP 132
+||||| |:|: |:|: |||||::|:: |:|: |||||:|
Db 4 MLNIALRAAKSAGELI---FRSIFRLDIVSVNEKDAKDYTEVDRAAEQTIVAALKAYP 60
+||||| |:|: |:|: |||||::|:: |:|: |||||:|
QY 133 THAIFGEENGW-KCAENSADFWVWLDPIDGKTSGFITGKPLFGTLIALHNGKVPVIGVIDQ 191
+||||| |:|: |:|: |||||::|:: |:|: |||||:|
Db 61 THAIMGEEGCFIEGCGEADYILWIDPLDGTNFIHGVPFAVSTACKYKGRLEHAYVLD 120
+||||| |:|: |:|: |||||::|:: |:|: |||||:|
QY 192 PILREKRWIGVKKQTTNQLAQEISVRSNLLAQAYLYTTSPLH-FEADAEDAFIR----V 245
+||||| |:|: |:|: |||||::|:: |:|: |||||:|
Db 121 PVKQEEFTASRKKAAALNFKRLKVSCKKSLFVALLAGTGTFPPRDNQIDNLDNLMFRSLV 180
+||||| |:|: |:|: |||||::|:: |:|: |||||:|
QY 246 RKKVAVPLYGCTNYAYALLASGPDVIVESGLKPYDFLSLVPVIECAGGSITDWRG---- 301
+||||| |:|: |:|: |||||::|:: |:|: |||||:|
Db 181 GUTAGIRKAGAASTDIAYVAGKYDAFWKPEGISPDMAAGALLVQFAGGLVSDFTGSHEF 2
```


Db 183 TREWRTRAYG-DFWPYMMVAEGSVDLCAEPESLMDMAAIIIVTEAGGFTGLDG--- 238

Qy 305 HWPVTAESRPTSENVVAAGDARVHKALDALRWR 338

Db 239 -----HPGPHSGN-AAASNGRLHDELLGYLNQR 265

RESULT 15

P95189

ID ID P95189 PRELIMINARY; PRQ; 260 AA.

AC P95189.

01-MAR-1997 (TEMUBLEL. 03, Last sequence update)
01-DEC-2001 (TEMUBLEL. 19, Last annotation update)
HYPOTHEICAL 27.7 KDA PROTEIN.
RV31137 OR MTCY03A2.21G.
Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxId=1773;
[1]
RN SEQUENCE FROM N.A.
RP STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
RL ENBL: Z83867. CAB06277.1. -

DR	infectio; IFK000760; inositol_P.
DR	Pfam; PF00459; inositol_P; 2.
DR	ProSITE; PS00129; IMP_1; 1.
KW	Hypothetical protein; complete proteome.
SQ	SEQUENCE 260 AA; 27 93 MW; CBCD2C2AA3E01518 CRC64;

Query Match	15.8%;	Score 273.5;	DB 16;	Length 260;
Best Local Similarity	32.5%;	Pred. No. 2.8e-14;		
Matches 88;	Conservative 42;	Mismatches 104;	Indels 37;	Gaps

QY	81	V A Q R A A D A A G E V L R - K Y F K Q R V E I I P K E D H S P V T I A D R E A E A M V S V I L K S F P T H A I F G E 130	
	:	: :	
	:	: :	
Db	8	L A L A D A R A D E I T E R V R C A I J L R L D T P K P T I T D A D R A V E S D V R Q T L G R D R P G D G V L G E 67	
OY	140	E N G W R C A E N S A D F V - - - W I L D I P D I G T K S F T G T G P L F C T L I A L L H N G K P V T G V I D O P I L R E 190	

[illegible]

Search completed: May 30, 2002, 17:16:34
Job time: 210 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 30, 2002, 17:04:58 ; Search time 32.63 Seconds
(without alignments)
1150.565 Million cell updates/sec

Title: US-09-686-522A-14
Perfect score: 1730
Sequence: 1 MLSSSSPHSDTSPPGLAS.....VVAAGDARVHKALDALRWR 338

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 74/574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_032802.*

- 1: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.*
- 2: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.*
- 3: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.*
- 4: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.*
- 5: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.*
- 6: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.*
- 7: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.*
- 8: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.*
- 9: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.*
- 10: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.*
- 11: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.*
- 12: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.*
- 13: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.*
- 14: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.*
- 15: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.*
- 16: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.*
- 17: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.*
- 18: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.*
- 19: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.*
- 20: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.*
- 21: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.*
- 22: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1730	100.0	338	21	AA1980.DAT.*
2	1294.5	74.8	324	21	AA1981.DAT.*
3	1067	61.7	295	21	AA1982.DAT.*
4	310.5	17.9	260	22	AA1983.DAT.*
5	310.5	17.9	260	22	AA1984.DAT.*
6	310.5	17.9	260	22	AA1985.DAT.*
7	286	16.5	265	19	AA1986.DAT.*
8	279	16.1	264	19	AA1987.DAT.*
9	267.5	15.5	271	21	AA1988.DAT.*
10	260	15.0	361	21	AA1989.DAT.*
11	260	15.0	371	21	AA1990.DAT.*

12	260	15.0	400	21	AA1991.DAT.*
13	255.5	14.8	267	20	AA1992.DAT.*
14	253.5	14.7	371	21	AA1993.DAT.*
15	253.5	14.7	400	21	AA1994.DAT.*
16	251.5	14.5	361	21	AA1995.DAT.*
17	244.5	14.1	266	22	AA1996.DAT.*
18	237.5	13.7	267	21	AA1997.DAT.*
19	236.5	13.7	333	21	AA1998.DAT.*
20	224	12.9	339	20	AA1999.DAT.*
21	221	12.8	168	21	AA2000.DAT.*
22	212	12.3	249	21	AA2001.DAT.*
23	212	12.3	284	22	AA2002.DAT.*
24	210.5	12.2	273	22	AA2003.DAT.*
25	210.5	12.2	288	22	AA2004.DAT.*
26	208.5	12.1	291	22	AA2005.DAT.*
27	208.5	12.1	291	22	AA2006.DAT.*
28	207	12.0	254	22	AA2007.DAT.*
29	207	12.0	397	21	AA2008.DAT.*
30	204	11.8	264	22	AA2009.DAT.*
31	201	11.6	278	22	AA2010.DAT.*
32	200	11.6	284	22	AA2011.DAT.*
33	199.5	11.5	270	15	AA2012.DAT.*
34	196.5	11.4	288	19	AA2013.DAT.*
35	195.5	11.3	253	22	AA2014.DAT.*
36	194.5	11.2	596	22	AA2015.DAT.*
37	193.5	11.2	277	15	AA2016.DAT.*
38	193.5	11.2	299	22	AA2017.DAT.*
39	188.5	10.9	325	20	AA2018.DAT.*
40	187.5	10.8	206	22	AA2019.DAT.*
41	185	10.7	406	21	AA2020.DAT.*
42	185	10.7	407	21	AA2021.DAT.*
43	185	10.7	416	21	AA2022.DAT.*
44	181	10.5	406	21	AA2023.DAT.*
45	181	10.5	407	21	AA2024.DAT.*

ALIGNMENTS

RESULT 1
AA1980.DAT.*
ID AA1980.DAT.* standard; Protein: 338 AA.
XX
AC AA1980.DAT.*
XX
DT 21-FEB-2000 (first entry)
XX
DE Corn extragenic suppressor protein.
XX
DE DE
XX
KW Clone cdt2c.pk003 b20; corn extragenic suppressor protein;
KW phytic acid biosynthetic enzyme; transformation; chimeric gene; antibody;
KW free phosphate; phytic acid level; transgenic plant; enzyme detection;
KW animal feed.
XX
OS zea mays.
XX
FH Key Location/Qualifiers
FT Misc-difference 11 /note= "Encoded by GCC"
FT FT
FT Misc-difference 94 /note= "Encoded by ANG"
FT FT
XX WO9955882-A1.
XX
PD 04-NOV-1999.
XX
PF Corynebacterium prote
XX Corynebacterium gl
XX Corynebacterium gl
XX Aquifex pyrophilus
XX Aquifex VF-5 alkali
XX Arabidopsis thalia
XX Arabidopsis thalia
XX Arabidopsis thalia

PI Cahoon RE, Hitz WD, Thorpe CJ, Tingey SV;

XX WPI: 2000-072181/06.
DR NP-SDB; AAZ29165.
XX
PT Novel phytic acid biosynthetic enzymes used to alter the level of the
PT enzyme in transgenic plants -
XX
XX Claim 6; Fig 2; 60pp; English.
PS
CC The present sequence is corn extragenic suppressor protein encoded by the
CC cDNA insert in clone cdt2c.pk003.b20. This belongs to the inositol
CC monophosphate family of proteins. The cDNA clone was isolated from corn
CC developing tassel. The level of expression of this phytic acid
CC biosynthetic enzyme can be altered by transforming a host cell with the
CC chimeric gene comprising the corresponding cDNA. This is done to increase
CC levels of free phosphate and decrease phytic acid levels in developing
CC seeds or in the construction of transgenic plants with altered levels of
CC the enzyme. The phytic acid biosynthetic enzymes are used to raise
CC antibodies, for enzyme detection in cells and cell extracts. This is
CC commercially used for animal feed.

Sequence 338 AA;

```

Query Match      100.0%; Score 1730; DB 21; Length 338;
Best Local Similarity 100.0%; Pred. No. 1e-169;
Matches 338; Conservative 0; Mismatches 0; Indels 0; Gaps 0
```

Qy	1	MLSSSSTHSDTSPFPGLASANPNRSLRLRAASPVSSAVLSAGSQPMSTVRASTAA	60
Dd	1	mlssssthsdstspfpglasanpnrsllrlraaspvssavilsagsgqpmstvrastaa	60
Qy	61	GACGRRAAAGELATERLVEVAQRADAAGADEVLRKYFRQRYEIDKEDHSPTYTIADRAE	120
Dd	61	gaagraaavgelaterlvevaqradaagaevlrkyfrqvyeidkedsptviadreae	120
Qy	121	EAMVSVLKSPTTHAIFGEENGWRCAEANSDFVWLDPIDGTSKITCKPLFGTLIALHL	180
Dd	121	eamvsvlksptthaifgeengwrcaeansdfvwlpidgtcksitgkplfgtliailh	180
Qy	181	NGKPVGIVGDIPILRRRWIGVDGQTTLNGOFISVRSCNLLAQAYLYTTSPHLFEADAED	240
Dd	181	ngkpvgivgdipilrrerwigvdgqttingqeisvrscnllaqaaylyttsphlfeadaed	240
Qy	241	AFIRVRNKVKPLYGCDCYAYALLASGFVDIVVESGLKPYDFLSLVVPVIEGAGGSITDWR	300
Dd	241	aafirvnkvplygcocyayallasgfvdvvvesylkpydfslsvvpiegagsitdwr	300
Qy	301	GDKUHWPTAESRPTSFNVVAAGDARVHKALDALRWR	338
Dd	301	gdkihwpvtaesrptsfnvvaagdardvhkealdalrwr	338

RESULT 2
AAV44233
ID AAY44233 standard; Protein: 324 AA.
XX
XX AC AAY44233;
XX
XX DT
XX DE 21-FEB-2000 (first entry)
XX DE Partial wheat extragenic suppressor protein.
XX KW Clone wlm0.pk0010.f6; wheat extragenic suppressor protein;
KW KW phytic acid biosynthetic enzyme; transformation; chimeric gene; antibody;
KW KW free phosphate; phytic acid level; transgenic plant; enzyme detection;
KW KW animal feed.
XX
OS Triticum aestivum.
XX
PN WO9955802-A1.
XX
PD 04-NOV-1999.

QY 309 TAESRPTSFNVAAAGDA-----RVHKEALDALR 336
 Db 241 -----ggdavatngilhdetldrik 260

RESULT 5
 AAB79381
 ID AAB79381 standard; Protein; 260 AA.
 AC AAB79381;
 XX
 DT 30-APR-2001 (first entry)
 DE
 DE Corynebacterium glutamicum SMP protein sequence SEQ ID NO:278.
 XX
 KW Corynebacterium glutamicum; carbon metabolism and energy production;
 KW SMP protein; sugar metabolism and oxidative phosphorylation protein;
 KW fine chemical production; organic acid; proteinogenic amino acid;
 KW nonproteinogenic amino acid; purine base; pyrimidine base; nucleoside;
 KW nucleotide; lipid; saturated fatty acid; unsaturated fatty acid; diol;
 KW carbohydrate; aromatic compound; vitamin; cofactor; polyketide; enzyme;
 diagnosis; Corynebacterium diphtheriae; evolutionary study.
 XX
 OS Corynebacterium glutamicum.

XX
 PN W0200100844-A2.
 XX
 PD 04-JAN-2001.
 XX
 PF 23-JUN-2000; 2000WO-1B00943.
 XX
 PR 25-JUN-1999; 99US-0141031.
 PR 08-JUL-1999; 99DE-1031412.
 PR 08-JUL-1999; 99DE-1031413.
 PR 08-JUL-1999; 99DE-1031419.
 PR 08-JUL-1999; 99DE-1031420.
 PR 08-JUL-1999; 99DE-1031424.
 PR 08-JUL-1999; 99DE-1031428.
 PR 08-JUL-1999; 99DE-1031431.
 PR 08-JUL-1999; 99DE-1031433.
 PR 08-JUL-1999; 99DE-1031434.
 PR 08-JUL-1999; 99DE-1031510.
 PR 08-JUL-1999; 99DE-1031562.
 PR 08-JUL-1999; 99DE-1031634.
 PR 09-JUL-1999; 99DE-1032227.
 PR 09-JUL-1999; 99DE-1032230.
 PR 09-JUL-1999; 99US-0143208.
 PR 14-JUL-1999; 99DE-1032924.
 PR 14-JUL-1999; 99DE-1032973.
 PR 14-JUL-1999; 99DT-1033005.
 PR 27-AUG-1999; 99DE-1040765.
 PR 31-AUG-1999; 99US-0151572.
 PR 03-SEP-1999; 99DE-1042076.
 PR 03-SEP-1999; 99DE-1042079.
 PR 03-SEP-1999; 99DE-1042086.
 PR 03-SEP-1999; 99DE-1042087.
 PR 03-SEP-1999; 99DE-1042088.
 PR 03-SEP-1999; 99DE-1042095.
 PR 03-SEP-1999; 99DE-1042123.
 PR 03-SEP-1999; 99DE-1042125.
 XX
 PA (BADI) BASF AG.
 XX
 XX Pompejus M, Kroeger B, Schroeder H, Zelder O, Habermayer G;
 WPI: 2001-061975/07.
 DR N-PSDB; AAB71498.
 XX
 XX New isolated Corynebacterium glutamicum nucleic acid encoding a sugar
 PT metabolism and oxidative phosphorylation protein for production or
 PT modulation of production of fine chemicals e.g. amino acids,
 PT -carbohydrates or enzymes -

XX Claim 20; Page 537-538; 1246pp; English.
 XX
 CC AAF71360 to AAF71750 encode the Corynebacterium glutamicum sugar
 CC metabolism and oxidative phosphorylation (SMP) proteins given in
 CC AAB79243 to AAB 79633 which are involved in carbon metabolism and
 CC energy production. The C. glutamicum SMP gene can be used in vectors
 CC (II) for expression in host cells and production or modulation of
 CC production of fine chemicals, such as, an organic acid, a proteinogenic
 CC or nonproteinogenic amino acid (preferred), a purine or pyrimidine base,
 CC a nucleoside, a nucleotide, a lipid, a saturated or unsaturated fatty
 CC acid, a diol, a carbohydrate, an aromatic compound, a vitamin, a
 CC cofactor, a polyketide, or an enzyme. The presence of (I) or SMP proteins
 CC (II) encoded by them as used for diagnosing the presence or activity of
 CC Corynebacterium diphtheriae in a subject. (I), (II), (III) or host cells
 CC containing them are used to map genomes of organisms related to
 CC C. glutamicum, identify and localize C. glutamicum sequences of interest,
 CC in evolutionary studies, in determining SMP protein regions required
 CC for function, in modulating SMP protein activity, in modulating the
 CC metabolism of sugars, and in modulating high-energy molecule production
 CC in a cell (i.e. ATP, NADPH).
 XX
 SQ Sequence 260 AA;

Query Match 17.9%; Score 310.5; DB 22; Length 260;
 Best Local Similarity 33.3%; Pred. No. 1.8e-23;
 Matches 91; Conservative 42; Mismatches 101; Indels 39; Gaps 11;

QY 81 VAQRADAAGEV-LRKYFRQVEIIDKEDHSPVTTADREAEMVSVLKSPFTHAIFGE 139
 Db 10 lalaeladstldrfeasdlvskpdmtpvsdadlateealrekiatarpadsilge 69
 QY 140 ENGWRCAENSADF---VWVLDPIIDGTGKSPITGKPIFGTLIALLHNGKPVIGVTDQPIRE 196
 Db 70 efg-----gdvefsgrqwiidpidgtknyrgvpwattialldngkpvagwisapalar 124
 QY 197 RWIGVDGKQT--TLNG---QEISVRSCNLLAQAYL-YTTSPLFEADAEDAFIRVANKV- 249
 Db 125 rwwasegagawrtfngsprklsvsqsklddaslsfsslsqgwaerldrdqfvsldttw 184
 QY 250 KVPYGCDCYAYALLASGFVDIVVESGLKPYDFLSLVPVIEGAGGSITDWRG-DKLHWPV 308
 Db 185 rlrqyg-dffsycivaegavdiaaepvslwdlapisilvteaggkftslagvvgph--- 240
 QY 309 TAESRPTSFNVAAAGDA-- KVIKEALDALR 436
 Db 241 -----ggdavatngilhdetldrik 260

RESULT 6
 AAB79382
 ID AAB79382 standard; Protein; 260 AA.
 AC AAB79382;
 XX
 DT 30-APR-2001 (first entry)
 DE
 DE Corynebacterium glutamicum SMP protein sequence SEQ ID NO:280.
 XX
 KW Corynebacterium glutamicum; carbon metabolism and energy production;
 KW SMP protein; sugar metabolism and oxidative phosphorylation protein;
 KW fine chemical production; organic acid; proteinogenic amino acid;
 KW nonproteinogenic amino acid; purine base; pyrimidine base; nucleoside;
 KW nucleotide; lipid; saturated fatty acid; unsaturated fatty acid; diol;
 KW carbohydrate; aromatic compound; vitamin; cofactor; polyketide; enzyme;
 diagnosis; Corynebacterium diphtheriae; evolutionary study.
 XX
 OS Corynebacterium glutamicum.
 XX
 PN W0200100844-A2.
 XX
 PD 04-JAN-2001.


```

Best Local Similarity 29.8%, Pred. No. 6.4e-21;
Matches 78; Conservative 53; Mismatches 105; Indels 26; Gaps 8;

QY 76 ERLVEVAQRAADAAGEVLRYKYF-RORVE-IIDKEDHSPVTIADREAEAMVSILKSFPT 133
   |:|:|:|:| |:|:|:|:| |:|:|:|:| |:|:|:|:| |:|:|:|:| |:|:|:|:| |:|:|:|:|
Db 5 ekilevamaalaggyvkenfgkiklenieekgekdfvsvydkteerikelilkffpd 64

QY 134 HATFGEENWRCAENSAFVWVLDPIDGTSFTGKPLFGTLTALLHNGKPVIGVIDQPI 193
   |:|:|:|:| |:|:|:|:| |:|:|:|:| |:|:|:|:| |:|:|:|:| |:|:|:|:| |:|:|:|:|
Db 65 hevvgdeekdegk--spykwfldpdktnykgfpifavsvglkenepivgavilpy 122

QY 194 LRERWIGVGKQTTLNGOEISVRSCNLLAQAYLYTTSPHLFEADA-----EDAFIRVR 246
   |:|:|:|:| |:|:|:|:| |:|:|:|:| |:|:|:|:| |:|:|:|:| |:|:|:|:| |:|:|:|:|
Db 123 fdilywaskrgaykogerisvkergeklkhaavvygfpssrdislylnvfkevfyev- 181

QY 247 NKVKVPLYGCCYAYALLASGFVDIVVSGELKPYDFLSLVPVIEGAGGSTIDWRGDKLHW 306
   |:|:|:|:| |:|:|:|:| |:|:|:|:| |:|:|:|:| |:|:|:|:| |:|:|:|:| |:|:|:|:|
Db 182 gsvrrp--gaaavdicmlaegifdgmmefemkpwditaglvilkeaggyft-lkgd---- 234

~ 307 PVTAESRPSTSFNVVAGDARVH 328
   |:|:|:|:| |:|:|:|:| |:|:|:|:| |:|:|:|:| |:|:|:|:| |:|:|:|:| |:|:|:|:|
235 -----pfgisdlagnrmilh 249

RESULT 8
AAW42388
ID AAW42388 standard; Protein; 264 AA.
XX AC AAW42388;
XX DT
XX DE
XX DE Aquifex VF-5 alkaline phosphatase (34A1A).
XX KW Alkaline phosphatase; thermostable enzyme; thermophilic bacterium;
XX KW food; detergent; baking.
XX OS Aquifex sp. strain VF-5.
XX PN WO9748416-A1.
XX PD 24-DEC-1997.
XX PF 19-JUN-1997; 97WO-US10784.
XX PR 19-JUN-1996; 96US-0033752.
XX PA (RECO-) RECOMBINANT BIOCATALYSIS INC.

Bylina E, Lee E, Mathur EJ;
WPI; 1998-062851/06.
N-PSDB; AAV03309.

Thermostable phosphatase(s) - useful in pharmaceutical, food,
detergent, and baking industries

Claim 10; Fig 9; 128pp; English.

This protein comprises a mature thermostable alkaline phosphatase
(34A1A) of Aquifex VF5, a marine bacterium that grows optimally
at 85-90 degC and pH 6.8. Its amino acid sequence, deduced from
an isolated polynucleotide (see AAV03309), shows 57% identity to
Escherichia coli suppressor protein subh. The invention relates to
claimed polynucleotides (see AAV03301-09 and AAV03310-20) coding for
claimed thermostable phosphatases (see AAW42380-95). Vector and host
cells are used to produce the enzymes, which can be used in a
claimed method to hydrolyse phosphate bonds. They can also be used
in enzyme labelling processes, in recombinant DNA techniques, in
ELISA, in enzyme linked gene probes, in research applications, for
removing 5' phosphates in polynucleotides prior to end labelling,
and in the pharmaceutical, food, detergent, and baking industries.

```

PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 21-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142154.
PR 06-JUL-1999; 99US-0142055.
PR 08-JUL-1999; 99US-0142390.
PR 09-JUL-1999; 99US-0142803.
PR 12-JUL-1999; 99US-0142920.
PR 13-JUL-1999; 99US-0142977.
PR 14-JUL-1999; 99US-0143542.
PR 15-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.

PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 31-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.

```
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match      15.5%; Score 267.5; DB 21; Length 271;
Best Local Similarity 30.1%; Pred. No. 5.4e-19;
Matches 81; Conservative 50; Mismatches 121; Indels 17; Gaps 7;

QY 76 ERLVEAQRADAGEVLKRFQRVEIIDKHDSPTVIADREAEAMVSVILKSFPTHA 135
Db 8 gllaaaiadaakagqirkyfctkvhvksqvdltvtdtgcceelvfnhklqifpnhk 67
136 IFGEEN--GWRCAENSADFVWVLDPIDGTGKSFITGKPLFGTLIALLLHNGKPVIGVIDOPI 193
Db 58 figeettaafgvteltdtdeptwivdpldttnfvhgfpcvcsigltikgvpvgvynpi 127
QY 194 LHERIGYDVGKOTTLNGQEI5VRSNLLAQAYLYT-TSPHLFEADAEDAFIRVRN--KV 249
Db 128 meeltfgvgkgafngkrikvsaqsqselltallvteagtkrdkatlddttrinsilckv 187
QY 250 K-VPLYGDCDYAYALLASGFVDIVVESGL-RPYDPLSLVPVIEGAGGSITDMRGDKLHWP 307
Db 188 rslrmsgscaldlcqvagcrvdfyelfgfgpwwdiaaqlvivkeagglifdpsgkdl--d 245
QY 308 VTAESRPTSFNVVAGDARVKEALDALR 336
Db 246 itsqr-----iaasnaslkeltaeair 267

RESULT 10
AAG11457
ID AAG11457 standard; Protein; 361 AA.
XX
AC AAG11457;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 10179.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
hybridisation assay; genetic mapping; gene expression control; promoter;
termination sequence.
..X
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 29-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR -23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 01-JUN-1999; 99US-0136782.
PR 03-JUN-1999; 99US-0137222.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140154.
PR 23-JUN-1999; 99US-0140154.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
```


PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134321.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 24-JUN-1999; 99US-0140823.
PR 28-JUN-1999; 99US-0140991.
PR 29-JUN-1999; 99US-0141287.
PR 30-JUN-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 01-JUL-1999; 99US-0142055.
PR 02-JUL-1999; 99US-0142239.
PR 06-JUL-1999; 99US-0142239.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.

PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 10-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.

```
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 13-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
PR 29-OCT-1999; 99US-0162142.

Query Match 15.0%; Score 260; DB 21; Length 371;
Best Local Similarity 26.9%; Pred. No. 5.1e-18;
Matches 92; Conservative 54; Mismatches 152; Indels 44; Gaps 10;

QY 13 SPFGGLASNPDRSLLRLRAASPYSSAVLSASGQPMSTYRASFAGAGRRRAAVGE 72
Db 27 nptsg-rtnrtftrctrlilnsfksktrlqt--kavisevsdqtrprig--aktgt 81
QY 73 LATERLVEAORAAAGAEVLKRYFRQVEIIDKDHSPVTIADREAEAMVSVILKSFP 132
Db 82 isphallevelaakgaeuvmeavknpritykglsldivtdtdkaseaaillevvkfnfs 141
QY 133 THAIFGEENGWRCASADFWVLLDPIDGTSKPTITGKPLFGTLIALHNGKP----VIGV 188
Db 142 dhiilgeegg-iigdsdssylwcidpldgttnfahgypsfaevsvgvlvrgnpaaasvvef 200
QY 189 IDQPILR--ERWTGVDGKQTTLLNGQEISVRSCNLLAQAYLYT-----TSPHLFE 235
Db 201 vggpmcrntrftsagaggalngqkhlvsktdaverallitfgyehddawstnnmelfk 260
QY 236 ADAEDAFIRVRNKVKPLXGDCDYAYALLASGVDIWSGLKPYDFLSLPVIEGAGGS 295
Db 261 -----eftdvsrgvr--rigaaavdmchvalgiaeswyeyrlkpwdmaagvlliveeagga 313
QY 296 ITDWRGD-----LHWPTAESRPTSFVVAAG 323
Db 314 vtrmdggktsvdfsvlsvngvhlphklleriapatenlkskg 355

RESULT 12
AAG11455
ID AAG11455 standard; Protein; 400 AA.
XX
AC AAG11455;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 10177.
XX
```

```
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
OS Arabidopsis thaliana.
PN EP1033405-A2.
XX
XX
PD 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
```

Protein identification; signal transduction pathway; metabolic pathway;
hybridisation assay; genetic mapping; gene expression control; promoter;
termination sequence.

Arabidopsis thaliana.

EP1033405-A2.

06-SEP-2000.

25-FEB-2000; 2000EP-0301439.

25-FEB-1999; 99US-0121825.

05-MAR-1999; 99US-0123180.

09-MAR-1999; 99US-0123548.

23-MAR-1999; 99US-0125788.

25-MAR-1999; 99US-0126264.

29-MAR-1999; 99US-0126785.

01-APR-1999; 99US-0127462.

06-APR-1999; 99US-0128234.

08-APR-1999; 99US-0128714.

16-APR-1999; 99US-0129845.

19-APR-1999; 99US-0130077.

21-APR-1999; 99US-0130449.

23-APR-1999; 99US-0130510.

23-APR-1999; 99US-0130891.

28-APR-1999; 99US-0131449.

30-APR-1999; 99US-0132048.

30-APR-1999; 99US-0132407.

04-MAY-1999; 99US-0132484.

05-MAY-1999; 99US-0132485.

06-MAY-1999; 99US-0132486.

06-MAY-1999; 99US-0132487.

07-MAY-1999; 99US-0132863.

11-MAY-1999; 99US-0134256.

14-MAY-1999; 99US-0134218.

14-MAY-1999; 99US-0134219.

14-MAY-1999; 99US-0134221.

14-MAY-1999; 99US-0134370.

18-MAY-1999; 99US-0134768.

19-MAY-1999; 99US-0134941.

20-MAY-1999; 99US-0135124.

21-MAY-1999; 99US-0135353.

24-MAY-1999; 99US-0135629.

25-MAY-1999; 99US-0136021.

27-MAY-1999; 99US-0136392.

Db 343 vtrmdgkfvdrsvlvsnvgvlhpklleriapatenlkskg 384

RESULT 13

AAW97883
ID AAW97883 standard; Protein; 267 AA.

XX
AC AAW97883;

DT 07-JUN-1999 (first entry)

DE Maize myo-inositol monophosphatase-3.

XX Myo-inositol monophosphatase-3; maize; corn; phytate;
KW phytic acid; transgenic plant; animal nutrition; feedstuff; food.

XX
OS Zea mays.

PN WO9905298-A1.

XX 04-FEB-1999.

XX 17-JUL-1998; 98WO-US14657.

XX 18-MAY-1998; 98US-0085852.

PR 22-JUL-1997; 97US-0053371.

PR 28-JUL-1997; 97US-0053944.

PR 08-AUG-1997; 97US-0055526.

PR 11-AUG-1997; 97US-0055446.

XX (PION-) PIONEER HI-BRED INT INC.

XX Beach LR, Bowen BA, Martino-Catt SJ, Wang H, Wang X;

XX WPI; 1999-142948/12.

DR N-PSDB; AAX24412.

XX New polynucleotides controlling phytate metabolism in plants -
PT useful for improving the nutritional content of plants, by enhancing
PT levels of non-phytate phosphorus, and reducing phytate levels

XX Claim 1a; Page 81-82; 86pp; English.

XX This is the amino acid sequence of maize myo-inositol
CC monophosphatase-3, an enzyme involved in the metabolism of
CC phytate. cDNA (see AAX24410) encoding the enzyme was isolated from
CC a maize immature ear cDNA library. Polynucleotides (see AAX24400,
CC AAX24403, AAX24407 and AAX24410-12) encoding maize
CC phosphatidylinositol-3-kinase (see AAW97880), myo-inositol
CC 1,3,4-triphosphate 5/6-kinase (see AAW97881), myo-inositol
CC 1-phosphate synthase and myo-inositol monophosphatase-3, all
CC enzymes involved in phytate metabolism, are claimed. The
CC invention relates to the use of such genes to reduce the
CC levels of phytate, and/or increase the levels of non-phytate
CC phosphorus, in plants used for food or feed. The genes are
CC especially used to improve the nutritional content of plants such
CC as corn and soybean. Transgenic plants, and seed produced by them,
CC are claimed.

XX Sequence 267 AA;

Query Match 14.8%; Score 255.5; DB 20; Length 267;

Best Local Similarity 30.3%; Pred. No. 9e-18;

Matches 73; Conservative 41; Mismatches 112; Indels 15; Gaps 5;

Qy 73 LATERLVEAQAADAAGEVLRYKRYRVEITIDKDESPVTIADREAEAMVSVILKSFP 132

Db 1 mseeqlavaveaakageiirkyfqtknvqhkgqvdltetdacedlfnhkrhfp 60

Qy 133 THAIFGENG---WRCAENSADFVWVLDPTDGTGKSFITGKPLFTGLALLHNGKPVIGVI 189

Db 61 dhkfigeeesaaalgatdltdptdwidpdtgttnfhgfpvcvsgltigkptvgvv 120

Qy 190 DQPILRERWIGVDGKQTTLNGQEISVRSCLNLAQAYLYTTSPHLFEADAEDAFIRVENKV 249
Db 121 fnplmelftavrgkgaflngspikassqdelvkalivteagtnrdkttvd---dtnrri 177
Qy 250 KVPLYGCD---CVAYAL----LASGFVDIWEVSGL-KPYDFLSLVPVIEGAGGSITDWR 300
Db 178 nrlllykirsirmcgsialnmcgvacgrldicyeigfggpdwaaagavilqeagglvfdps 237
Qy 301 G 301
Db 238 g 238

RESULT 14

AAG39265

ID AAG39265 standard; Protein; 371 AA.

XX AAG39265;

XX 18-OCT-2000 (first entry)

XX Arabidopsis thaliana protein fragment SEQ ID NO: 48558.

KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.

XX Arabidopsis thaliana.

XX EP1033405-A2.

XX 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-01 825.

PR 05-MAR-1999; 99US-014 180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.

PR 28-APR-1999; 99US-0130891.

PR 30-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132048.

PR 04-MAY-1999; 99US-0132407.

PR 05-MAY-1999; 99US-0132484.

PR 06-MAY-1999; 99US-0132485.

PR 06-MAY-1999; 99US-0132486.

PR 07-MAY-1999; 99US-0132487.

PR 11-MAY-1999; 99US-0132863.

PR 14-MAY-1999; 99US-0134256.

PR 14-MAY-1999; 99US-0134218.

PR 14-MAY-1999; 99US-0134219.

PR 14-MAY-1999; 99US-0134221.

PR 18-MAY-1999; 99US-0134370.

PR 19-MAY-1999; 99US-0134768.

PR 20-MAY-1999; 99US-0134941.

PR 21-MAY-1999; 99US-0135124.

PR 24-MAY-1999; 99US-013529.

PR 25-MAY-1999; 99US-0135629.

PR 27-MAY-1999; 99US-0136021.

PR 28-MAY-1999; 99US-0136392.

PR 01-JUN-1999; 99US-0136782.

PR 03-JUN-1999; 99US-0137222.

PR 99US-0137528.

PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142820.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.

PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149829.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157965.
PR 07-OCT-1999; 99US-0157965.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0158293.
PR 13-OCT-1999; 99US-0159294.
PR 14-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159340.
PR 14-OCT-1999; 99US-0159341.
PR 14-OCT-1999; 99US-0159647.
PR 14-OCT-1999; 99US-0159648.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match

Best Local Similarity

14.7%
25.8%;Score 253.5; DB 21;
Pred. No. 2.4e-17;
Length 371;

Matches 93; Conservative 55; Mismatches 165; Indels 47; Gaps 10;	
Qy 1	MLSSSTHSDTSPFPGCLANP---NPRSLRLRLRAASPVSSAVLSASGRQP---MSTV 54
Db 6	ifsgnmalrshlprsslpignpigtvtrtfryrctrlilnsfksttrlqtkavlsev 65
Qy 55	RASFAAGAAAGRAAAGELATERLVEVAQRAADAAGEVLKRYFRQRVEIIDKEDHSPVTI 114
Db 66	sdqtryrig-aktttisphalleveelaaktgaevm'avnkprnietykgsdlvtd 123
Qy 115	ADRAEAMVSVILKSPPTHAIFGEENGWCAENSADFWVLDPIDGKTSFITGKPLFGT 174
Db 124	tdkaseailevknfshllegeeg-iiigdsstqylwciqldgttnfahgybsfav 182
Qy 175	LIALHNGKE---VICVIDQPL--RERWIGVDGKOTTLNGOEISVRSCNLLAQAYLYT 228
Db 183	svgilyrgnpaaasvvefvggpmcwntrtfsataggalncgkqihvskdaverallit 242
Qy 229	-----TSPHLFEADAEDAFIRNRKVKVPLYGCDYAYALLASGFVIVVESGL 277
o 243	gfgyehddawstnmelrk-----eftdvsrgvr--rlgaaavdmchvalgiaesweyrl 295
Qy 278	KPYDFLSLVPVIEGAGSITDWRCDK-----LHWPVTAESRPTSFNVVAAG 323
Db 296	kpwmdmaagvliveagavtrmdgkfsfdrsvlsvngvlhpklleriapatenlkskg 355
RESULT 15	
AAG39264	
XX	ID AAG39264 standard; Protein: 400 AA.
AC	AAG39264;
XX	
DT	18-OCT-2000 (first entry)
DE	Arabidopsis thaliana protein fragment SEQ ID NO: 48557.
XX	
KW	Protein identification; signal transduction pathway; metabolic pathway;
KW	hybridisation assay; genetic mapping; gene expression control; promoter;
KW	termination sequence.
XX	
OS	Arabidopsis thaliana.
XX	
PN	EP1033405-A2.
XX	
PD	06-SEP-2000.
XX	
FF	25-FEB-2000; 2000EP-0301439.
X	
rR	25-FEB-1999; 99US-0121825.
PR	05-MAR-1999; 99US-0123180.
PR	09-MAR-1999; 99US-0123548.
PR	23-MAR-1999; 99US-0125788.
PR	25-MAR-1999; 99US-0126264.
PR	29-MAR-1999; 99US-0126785.
PR	01-APR-1999; 99US-0127462.
PR	06-APR-1999; 99US-0128234.
PR	08-APR-1999; 99US-0128714.
PR	16-APR-1999; 99US-0129845.
PR	19-APR-1999; 99US-0130077.
PR	21-APR-1999; 99US-0130449.
PR	23-APR-1999; 99US-0130510.
PR	23-APR-1999; 99US-0130891.
PR	28-APR-1999; 99US-0131449.
PR	30-APR-1999; 99US-0132048.
PR	30-APR-1999; 99US-0132407.
PR	04-MAY-1999; 99US-0132484.
PR	05-MAY-1999; 99US-0132485.
PR	06-MAY-1999; 99US-0132486.
PR	06-MAY-1999; 99US-0132487.
PR	07-MAY-1999; 99US-0132863.
PR	11-MAY-1999; 99US-0134256.
PR	14-MAY-1999; 99US-0134218.
PR	
PR	14-MAY-1999; 99US-0134219.
PR	14-MAY-1999; 99US-0134221.
PR	14-MAY-1999; 99US-0134370.
PR	18-MAY-1999; 99US-0134768.
PR	19-MAY-1999; 99US-0134941.
PR	20-MAY-1999; 99US-0135124.
PR	21-MAY-1999; 99US-0135353.
PR	24-MAY-1999; 99US-0135629.
PR	25-MAY-1999; 99US-0136021.
PR	27-MAY-1999; 99US-0136392.
PR	28-MAY-1999; 99US-0136782.
PR	01-JUN-1999; 99US-0137222.
PR	03-JUN-1999; 99US-0137528.
PR	04-JUN-1999; 99US-0137502.
PR	07-JUN-1999; 99US-0137724.
PR	08-JUN-1999; 99US-0138094.
PR	10-JUN-1999; 99US-0138540.
PR	10-JUN-1999; 99US-0138847.
PR	14-JUN-1999; 99US-0139119.
PR	16-JUN-1999; 99US-0139452.
PR	16-JUN-1999; 99US-0139453.
PR	17-JUN-1999; 99US-0139492.
PR	18-JUN-1999; 99US-0139454.
PR	18-JUN-1999; 99US-0139455.
PR	18-JUN-1999; 99US-0139456.
PR	18-JUN-1999; 99US-0139457.
PR	18-JUN-1999; 99US-0139458.
PR	18-JUN-1999; 99US-0139459.
PR	18-JUN-1999; 99US-0139460.
PR	18-JUN-1999; 99US-0139461.
PR	18-JUN-1999; 99US-0139462.
PR	18-JUN-1999; 99US-0139463.
PR	18-JUN-1999; 99US-0139750.
PR	18-JUN-1999; 99US-0139763.
PR	21-JUN-1999; 99US-0139817.
PR	22-JUN-1999; 99US-0139899.
PR	23-JUN-1999; 99US-0140353.
PR	23-JUN-1999; 99US-0140354.
PR	24-JUN-1999; 99US-0140695.
PR	28-JUN-1999; 99US-0140823.
PR	29-JUN-1999; 99US-0140991.
PR	30-JUN-1999; 99US-0141287.
PR	01-JUL-1999; 99US-0141842.
PR	01-JUL-1999; 99US-0142154.
PR	02-JUL-1999; 99US-0142055.
PR	06-JUL-1999; 99US-0142490.
PR	08-JUL-1999; 99US-0142803.
PR	09-JUL-1999; 99US-0142920.
PR	12-JUL-1999; 99US-0142977.
PR	13-JUL-1999; 99US-0143542.
PR	14-JUL-1999; 99US-0143624.
PR	15-JUL-1999; 99US-0144005.
PR	16-JUL-1999; 99US-0144085.
PR	16-JUL-1999; 99US-0144086.
PR	19-JUL-1999; 99US-0144325.
PR	19-JUL-1999; 99US-0144331.
PR	19-JUL-1999; 99US-0144332.
PR	19-JUL-1999; 99US-0144333.
PR	19-JUL-1999; 99US-0144334.
PR	19-JUL-1999; 99US-0144335.
PR	20-JUL-1999; 99US-0144352.
PR	20-JUL-1999; 99US-0144632.
PR	20-JUL-1999; 99US-0144884.
PR	21-JUL-1999; 99US-0144814.
PR	21-JUL-1999; 99US-0145086.
PR	21-JUL-1999; 99US-0145088.
PR	22-JUL-1999; 99US-0145085.
PR	22-JUL-1999; 99US-0145087.
PR	22-JUL-1999; 99US-0145089.
PR	23-JUL-1999; 99US-0145192.
PR	23-JUL-1999; 99US-0145145.
PR	23-JUL-1999; 99US-0145218.
PR	23-JUL-1999; 99US-0145224.

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 17, 2002, 16:19:08 ; Search time 45 Seconds
(without alignments)
722.076 million cell updates/sec

Title: US-09-686-522C-14
Perfect score: 1730
Sequence: 1 MLSSSSSTHSDTSPFGLAS.....VVAAGDARVHKALDALRW 338

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR73:*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1020	59.0	286	2 T08560	hypothetical prote
2	525	30.3	262	2 AG3034	inositol monophosp
3	525	30.3	297	2 E98251	similar to micb177
4	483	27.9	308	2 F95293	probable inositol
5	477	27.6	278	2 C87528	inositol monophosp
6	463.5	26.8	272	2 AF3515	extragenic suppres
7	396.5	22.9	267	2 S74899	extragenic suppres
8	293	16.9	267	2 AC0353	probable inositol
9	292	16.9	271	2 C83169	extragenic suppres
10	283	16.4	255	2 T45317	monophosphatase (im
11	279	16.1	264	1 C70470	myo-inositol-1(or
12	273.5	15.8	260	2 B70646	probable monophosp
13	272	15.7	267	1 F64103	suppressor protein
14	265.5	15.3	268	2 T07795	inositol-1(or 4)-m
15	258	14.9	267	1 D65030	inositol-1(or 4)-m
16	258	14.9	267	2 G91053	extragenic suppres
17	257	14.9	267	2 A10824	extragenic suppres
18	254.5	14.7	270	2 AF2170	inositol monophosph
19	248	14.3	288	2 D82285	inositol monophosph
20	245	14.2	266	2 AG3580	inositol-1(or 4)-m
21	244.5	14.1	269	2 AG3145	inositol monophosph
22	244.5	14.1	276	2 D98142	sPCA protein (AF17
23	244	14.1	263	2 AH3292	inositol-1(or 4)-m
24	240	13.9	266	2 H98266	extragenic suppres
25	240	13.9	266	2 AH3017	extragenic suppres
26	239.5	13.8	266	2 E84292	extragenic suppres
27	238	13.8	265	2 T07800	inositol-1(or 4)-m
28	236.5	13.7	268	2 F97443	inositol monophosph
29	236.5	13.7	268	2 AH2661	inositol monophosph

30	235	13.6	270	2 T12205	inositol-1(or 4)-m
31	234	13.5	275	2 E82554	extragenic suppres
32	231.5	13.4	353	2 A86438	hypothetical prote
33	230	13.3	287	1 S76736	hypothetical prote
34	229.5	13.3	256	2 T50642	myo-inositol-1(or
35	229.5	13.3	275	2 AE3393	inositol monophosp
36	226.5	13.1	232	2 E72255	hypothetical prote
37	221.5	12.8	265	1 E69864	myo-inositol-1(or
38	221	12.8	266	2 D95885	probable inositol
39	221	12.8	272	2 T03436	accG protein - Agr
40	221	12.8	272	2 AG3245	arabinose phosphat
41	221	12.8	280	2 AB2875	exopolysaccharide
42	220	12.7	262	2 C97651	psbB protein [impo
43	220	12.7	294	2 G95861	probable inositol
44	218	12.6	257	2 AB1208	extragenic suppres
45	218	12.6	304	2 G97583	inositol monophosp

ALIGNMENTS

RESULT 1

T08560
hypothetical protein T22F8.20 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 17-Mar-2000
C:Accession: T08560
R:Bevan, M.; Zimmermann, W.; Gruenisen, A.; Wambutt, R.; Bancroft, I.; Meves, H.W.;
submitted to the Protein Sequence Database, May 1999
A:Reference number: Z16442
A:Accession: T08560
A:Molecule type: DNA
A:Residues: 1-286 <BEV>
A:Cross-references: EMBL:AL050351; GSPDB:GN00062; ATSP:T22F8.20
A:Experimental source: cultivar Columbia; BAC clone T22F8
C:Genetics:
A:Gene: ATSP:T22F8.20
A:Map position: 4
A:Introns: 58/1; 84/2; 116/1; 132/3; 174/2; 196/3; 226/3; 264/1
C:Superfamily: suppressor protein subB

Query Match	Best Local Similarity	Score	1020;	DB 2:	Length 286;
Matches 182;	Conservative 45;	Mismatches 35;	Indels 0;	Gaps 0;	
QY	76	ERLVEVQRAADAAGEVLRYRQVEI	DKEDSHSPVTIADREAEAMVSVILKSPFTHA	135	
Db	24	DRFAAVGNALADASGEVIRKVF	KKFDIVDKDMSPTTIADQMAEAMVSIIFQNLPSHA	83	
QY	136	IFGEENGWRCANSAFVWVLDPI	DGTSKSFITGKPLFGLTIALLLHNGRPVIGVQPILR	195	
Db	84	IYGEKGRCKEESADYVWVLDPI	DGTSKSFITGKPVFTIALLYKGPILGLIDQPILR	143	
QY	196	ERWIGVCKQTLLNGOETSVRSC	NLLAOAYLTTSPHLFEADAEDAFIRVKRVKVPYLG	255	
Db	144	ERWIGMGRRTKLNEDISTRCP	KLQAYLYTTSFHLFEAEAKYASVRDKVPLYG	203	
QY	256	CDYAYALLASGFVDIVVESGL	KPYDFLSLPVIEGAGSITDWRGDKLHWPTAESRPT	315	
Db	204	CDYAYALLASGFVDLVIESGL	KPYDFLALVPVIEGAGGITDWTGKRFLWEASSAVAT	263	
QY	316	SNVVAAGDARVHKALDALRW	337		
Db	264	SNVVAAGDSIHQALLESLEW	285		

RESULT 2

AG3034
inositol monophosphatase family protein [imported] - Agrobacterium tumefaciens (stra
C:Species: Agrobacterium tumefaciens
C>Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 01-Feb-2002
C:Accession: AG3034
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Wc

erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClellan, Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, S.E.W.

A:Title: The Genome of the Natural Genetic Engineer *Agrobacterium tumefaciens* C58.

A:Reference number: AB2577; PMID:11743193

A:Accession: AG3034

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-262 <KUR>

A:Cross-references: GB:AE008689; PIDN:AAU44693.1; PID:g17742322; GSPDB:GN00187

A:Experimental source: strain C58 (Dupont)

C:Genetics:

A:Gene: Atu3885

A:Map position: linear chromosome

C:Superfamily: suppressor protein subB

Query Match 30.3%; Score 525; DB 2; Length 262;

Best Local Similarity 42.7%; Pred. No. 7.2e-34;

Matches 112; Conservative 51; Mismatches 73; Indels 26; Gaps 7;

Qy 84 RAADAAGEVLRYRFRVEIDKED--HSPVTIADREAEAMVSVILKSFTHAIFGEEN 141

Db 14 KLADAASAETLPRFRFGIAVINKQGGYDPVTEGQAAETAIRALIEERFPOHGILGEEH 73

Qy 142 GWRCACEN---SADFVWVLDPIDGTSFKITGKPLFGLTIALHNGKPVIGVDOPIILRERW 198

Db 74 G-----NVGLDRDHIIWIDPIDGTRAFISGVPVWMTLIGFOSSGRATMGIMDQPTTKERY 128

Qy 199 IGVGDKQOTTLNG-----QEISVRSNLLAQAYLYTTSPLHFEADAEDAFIRVKNKVPLY 254

Db 129 F-ADGKAAYFGPDGEEKIRTRDCASLSDAVLTFTTHIFTAAEEKPLYEKVQDQVRLFRY 187

Qy 255 GDCYAYALLASGFVDIVVESGLKPYDFLSLPVIEGAGGSITDWRGDKLHWPVTAESRP 314

Db 188 GVDCYAYCLLAAGHVDLVIESGLKPYDVGALIPVIEQAGTWTWTDG-----GRP 237

Qy 315 -TSFNVAAGDARVHKEALDAL 335

Db 238 ENGRILAAGSKAVHEEALAIL 259

RESULT 3

E98251 Similar to mlcb1779.29 (probable monophosphatase gene) in *mycobacterium leprae* [imported]

C:Species: *Agrobacterium tumefaciens*

C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 15-Feb-2002

A:Accession: E98251

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-297 <KUR>

A:Cross-references: GB:AE007870; PIDN:AAK89535.1; PID:g15159414; GSPDB:GN00170;

C:Genetics:

A:Gene: AGR_L1926

A:Map position: linear chromosome

C:Superfamily: suppressor protein subB

Query Match 30.3%; Score 525; DB 2; Length 297;

Best Local Similarity 42.7%; Pred. No. 8.4e-34;

Matches 112; Conservative 51; Mismatches 73; Indels 26; Gaps 7;

Qy 84 RAADAAGEVLRYRFRVEIDKED--HSPVTIADREAEAMVSVILKSFTHAIFGEEN 141

Db 49 KLADAASAETLPRFRFGIAVINKQGGYDPVTEGQAAETAIRALIEERFPOHGILGEEH 108

Qy 142 GWRCACEN---SADFVWVLDPIDGTSFKITGKPLFGLTIALHNGKPVIGVDOPIILRERW 198

Db 109 G-----NVGLDRDHIIWIDPIDGTRAFISGVPVWMTLIGFOSSGRATMGIMDQPTTKERY 163

Qy 199 IGVGDKQOTTLNG-----QEISVRSNLLAQAYLYTTSPLHFEADAEDAFIRVKNKVPLY 254

Db 164 F-ADGKAAYFGPDGEEKIRTRDCASLSDAVLTFTTHIFTAAEEKPLYEKVQDQVRLFRY 222

Qy 255 GDCYAYALLASGFVDIVVESGLKPYDFLSLPVIEGAGGSITDWRGDKLHWPVTAESRP 314

Db 223 GVDCYAYCLLAAGHVDLVIESGLKPYDVGALIPVIEQAGTWTWTDG-----GRP 272

Qy 315 -TSFNVAAGDARVHKEALDAL 335

Db 273 ENGRILAAGSKAVHEEALAIL 294

RESULT 4

F95293

probable inositol monophosphatase [imported] - *Sinorhizobium meliloti* (strain 1021)

C:Species: *Sinorhizobium meliloti*

C:Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001

C:Accession: F95293

R:Barnett, M.J.; Fisher, R.F.; Jones, T.; Komp, C.; Abola, A.P.; Barloy-Hubler, F.;

; Kalman, S.; Keating, D.H.; Palm, C.; Peck, M.C.; Surzycki, R.; Wells, D.H.; Yeh,

Proc. Natl. Acad. Sci. U.S.A. 98, 9893-9898, 2001

A:Title: Nucleotide sequence and predicted functions of the entire *Sinorhizobium mel*

A:Reference number: A95262; MUID:21396509; PMID:11481432

A:Accession: F95293

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-308 <KUR>

A:Cross-references: GB:AE006469; PIDN:AAK64912.1; PID:g14523333; GSPDB:GN00165

A:Experimental source: strain 1021, megaplasmid pSymA

R:Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hub

pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R

L.; Hyman, R.W.; Jones, T.

Science 293, 668-672, 2001

A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lela;

hebaull, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh

A:Title: The composite genome of the legume symbiont *Sinorhizobium meliloti*.

A:Reference number: A96039; MUID:21368234; PMID:11474104

A:Contents: annotation

C:Genetics:

A:Gene: Sma0483

A:Genome: plasmid

Query Match 27.9%; Score 483; DB 2; Length 308;

Best Local Similarity 39.6%; Pred. No. 1.8e-30;

Matches 107; Conservative 38; Mismatches 103; Indels 22; Gaps 4;

Qy 70 VGEATERLVEAQRADAAGAEVLKRYRQVEIIDKEDHSPVTIADREAEAMVSVILK 129

Db 56 LGEFAS-----FAHDIADLAKQTSSAGVREPIAKSDASPVTTEDRAVEKLEERAD 110

Qy 130 SFPTHAIFGEENGWRCACENSAADFVWVLDPIDGTSFKITGKPLFGLTIALHNGKPVIGV 189

Db 111 HFPDGHVLGEFEAGELGN--EFVWVIDPIDGTAKFAVAGLPVYGLTSLTRGGTPIGLI 168

Qy 190 DQPILRERWIGDKQOTTLNGQEISVRSNLLAQAYLYTTSPLHFEADAEDAFIRVKNV 249

Db 169 DNPMTGDRMLGVSGQPTTLNNVPIRTASTATATATAFIANGNPDAFSPADKSRVESLRIT 228

Qy 250 KVPYLGDCYAYALLASGFVDIVVESGLKPYDFLSLPVIEGAGGSITDWRGDKI.HWPVT 309

Db 229 RWCYVGGSCIAVGRVADGSDVISIDGGLDPDYCYALVPVITGAGCITDWG----- 280

Qy 310 AESRPTSFN----VVAAGDARVHKEALDAL 335

Db 281 ---RPLTLNSGCLCATATDILLHRHVLIL 307

RESULT 5

C87528

inositol monophosphatase family protein [imported] - Caulobacter crescentus
C:Species: Caulobacter crescentus
C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 01-Mar-2002
C:Accession: C87528
R: Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon
B.; Laub, M.T.; Deboy, R.T.; Dodson, R.J.; Shaprio, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A:Title: Complete Genome Sequence of Caulobacter crescentus.
A:Reference number: A87249; MUID:21173698; PMID:11259647
A:Accession: C87528
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-278 <STO>
A:Cross-references: GB:AE005673; NID:g13423761; PIDN:AAK24223.1; GSPDB:GN00148
C:Genetics:
A:Gene: CG2252
C:Superfamily: suppressor protein subB

Query Match 27.6%; Score 477; DB 2; Length 278;
Best Local Similarity 40.8%; Pred. No. 4.6e-30;
Matches 116; Conservative 48; Mismatches 88; Indels 32; Gaps 9;
Qy 73 LATERLVEAQAAD---AAGEVLRKYFRQVRIIDK-----EDHS---PVTIADRE 118
Db 3 LSADRLAALDAFILDNLASADVILPLFRADHGLEKDGAGKLNLPRTDTHAAFDVPTADRG 62
Qy 119 AEEAMVSVILKSPPTTHAIFGEENGWCAEN--SADFVWLDPIDGKSFITGKPLFGTLI 176
Db 63 AEAAIRALIAQRFPHGVIGEEYG---EDRPAEAFVWLDPIDGTRAFIN .PLWTLLI 118
Qy 177 ALLHNGKPVTVGIDOPTLRERWIGVGDKQTTLLNGQE---ISVRSNLLAQAYLYTTSPH- 232
Db 119 GLRHEGRPVLSIGCPVYNEIFIGHAGGARLVSGGEARPIRVRECANINDAVIATTPDA 178
Qy 233 LFEADAEDAFIRVANKVKVPLYGCDYAYALLASGFVDIVVESGLKPYDFLSLVPIEGAG 292
Db 179 CFDAERGAWLQVRAAKLARLGCDAAYAMVAMGKMDV EAGLSWDLAAIPIIEGA 238
Qy 293 GGSITDHRGDKLHWPVTAESRPTSFNVVAAGDARVHKEALDALR 336
Db 239 GGMVTNMRGE-----PV-----GPGGGQGVISGDRRPLDEALVSLR 274

RESULT 6
AF3515
Extragenic suppressor protein subB [imported] - Brucella melitensis (strain 16M)
C:Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 15-Feb-2002
;Accession: AF3515
;DelVecchio, V.G.; Kapatal, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova,
; Mazur, M.; Goldsman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letess
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A:Title: The genome sequence of the facultative intracellular pathogen Brucella melitensis
A:Reference number: AD3252; PMID:11756688
A:Accession: AF3515
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-272 <KUR>
A:Cross-references: GB:AE008918; PIDN:AAL53289.1; PID:g17984172; GSPDB:GN00191
A:Experimental source: strain 16M
C:Genetics:
A:Gene: BMEI0048
A:Map position: II
C:Superfamily: suppressor protein subB

Query Match 26.88; Score 463.5; DB 2; Length 272;
Best Local Similarity 41.38; Pred. No. 5.2e-29;
Matches 116; Conservative 47; Mismatches 93; Indels 25; Gaps 9;
Qy 66 RAAAVGELATERLVEVA---QRAADAAGEVLRKYFRQVRIIDK---HSPVTIADREAE 120
Db 4 RAPCFGNRRTGLLIDKFAFFSEVAAAAAQTLP-FRQLTVDNKYSVGFDVPTADRAAE 62

Qy 121 EAMVSVILKSPPTTHAIFGEENGWCAEN--ADFVWLDPIDGKSFITGKPLFGTLIALL 179
Db 63 RAIRAVIGRTFPHYHILGEEYG---AENTDRSHVWIIDPVDGTRAFISGLPVMGTLVGLT 119
Qy 180 HNGKPVTVGIDOPTLRERWIGVGDKQTTLLN-----QGEISVRSNLLAQAYLYTTSPHLF 234
Db 120 VGDGARAQMSPPTGELFYS-DGDCAYLQDGDAPRCLCVKRNNAVLEDAITFTTTPALF 178
Qy 235 EADAEDAFIRVANKVKVPLYGCDYAYALLASGFVDIVVESGLKPYDFLSLVPIEGAGG 294
Db 179 KGDDRKAFLRLESARVLSRYGVDCYAFAMLAGGFVDIVVEAGLQTYDIAALIPTEQAGG 238
Qy 295 SITDHRGDKLHWPVTAESRPTSFNVVAAGDARVHKEALDAL 335
Db 239 VYTRRDGG-----PAEQ---GGDIVAAATPALHQAALDLL 270

RESULT 7
S74899
Extragenic suppressor - Synecocystis sp. (strain PCC 6803)
N:Alternate names: protein sll1959
C:Species: Synecocystis sp.
A:Variety: PCC 6803
C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 20-Jun-2000
C:Accession: S74899
R: Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miya, J.
Okumura, K.; Shimpo, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Y.
DNA Res. 3, 109-136, 1996
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synecoc.
s.
A:Reference number: S74322; MUID:97061201; PMID:8905231
A:Accession: S74899
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-267 <KAN>
A:Cross-references: EMBL:D90909; GB:AB001339; NID:g1652844; PIDN:BAAL7860.1; PID:gl
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C:Superfamily: suppressor protein subB

Query Match 22.9%; Score 396.5; DB 2; Length 267;
Best Local Similarity 36.9%; Pred. No. 9.4e-24;
Matches 101; Conservative 49; Mismatches 99; Indels 25; Gaps 10;
Qy 76 ERLVEVAQRAADAAGEVLRKYFR---QRVEIDKEDHSPVTIADREAEAMVSVILKSF 131
Db 6 EQLRLFAQQLAAVSGEILAIQYFKRSHLQGTCKIDQVS-AIVTQADEAEQAQAVDLIOAQF 64
Qy 132 PTHAIFGEENGWCAENSAHNVWLDPIDGKSFITGKPLFGTLIALLH-NCKPVIGVID 190
Db 65 PQDGVIREE-GKNJACKSG-YTWVLDPIDGTSFVGRGLPIFATLGLVDADMRPVIGIAH 122
Qy 191 QPILRERWIGVGDKQTTLLNG---QEISVRSNLLAQAYLYTTSPHLFEADA-EDAFIRV 245
Db 123 QPISGDRWQVQGEQSNVNGIPLVNPYKASEINLTA-ACIVSTTPLMFTTPVQOQKADI 181
Qy 246 RNKVKVPLYGCDYAYALLASGFVD---IVVESGLKPYDFLSLVPIEGAGSGITDWRGD 302
Db 182 YRCQRTAFGGDCFNFLSAASGTAMPPLVIVEADLNFYDFCALIPTILTGANTCYFTDQMGK 241
Qy 303 KLHWPVTAESRPTSFNVVAAGDARVHKEALDALR 336
Db 242 EL-----TPESTEVASPNKLHSEILAFLO 267

RESULT 8
AC0353
Probable inositol monophosphatase family protein YPO2899 [imported] - Yersinia pest
C:Species: Yersinia pestis
C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 09-Nov-2001
C:Accession: AC0353
R: Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice,
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan,

C:Genetics:
A:Gene: imp2
C:Superfamily: suppressor protein subh

Query Match 16.1%; Score 279; DB 1; Length 264;
Best Local Similarity 28.5%; Pred. No. 1.6e-14;
Matches 75; Conservative 54; Mismatches 106; Indels 28; Gaps 7;

QY 76 ERLVEVAQRAADAGAEVLKRYFR--QRVEIIDKEDHSPVTIADRAEAEAMVSVILKSPPT 133
DB 5 KYLEVAKIAALAGGQVLKENGKVKENIEKGKDFVSYVDKTSSEIRKEVILKFFPD 64
QY 134 HAIFGEGNRCACNS-ADFWVLDPIIDGTSFKITGKPLFGTLIALHNGKPVIGVIDOP 192
DB 65 HEVVGEMG---AEGSGSYRFIDPLDGTKNYNGFIFAVSVGLYKGEFIVGAVYLP 121
QY 193 ILRERIGVGDQKOTTLNGQEISVRSNCNLAQAYLYTTSPLFEADA-----EDAFIRV 245
DB 122 YFDKLYGAKGLGAYVNGKRIKVKDNESLKHAGVYVGFPSRRRDISIYLNIFKDVFEV 181
QY 246 RNKVKVPYLYGCDYAYALLAGFVDIVVESGLKPYDFLSVPVIEGAGGSITDWRGDKLH 305
DB 182 -GSMRRP--GAAAVDLCVMAEIGFDMFEKMPWDITA'LVLKBAAG----- 227
QY 306 WPVTAESRPTSFNVVAGDARVH 328
DB 228 -VYTLVGEPFGVSDIAGNKAH 249

RESULT 12
B70646
probable monophosphatase - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: B70646
R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998
A: Authors: Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A: Reference number: A70500; MUID: 98295987; PMID: 9634230
A: Accession: B70646
A: Status: preliminary; nucleic acid sequence not shown; translation not shown
A: Molecule type: DNA
A: Residues: 1-260 <COL>
A: Cross-references: GB:Z83867; GB:AL123456; NID:G3261695; PIDN: CAB06277.1; PID: g1781230
A: Experimental source: strain H37RV
C: Genetics:
C: Gene: RV3137
C: Superfamily: suppressor protein subh

Query Match 15.8%; Score 273.5; DB 2; Length 260;
Best Local Similarity 32.5%; Pred. No. 4.3e-14;
Matches 88; Conservative 42; Mismatches 104; Indels 37; Gaps 11;

QY 81 VAQRAADAGEVLR-KYFRQVEIIDKEDHSPVTIADRAEAEAMVSVILKSPPTAIFCE 139
DB 8 LALALADRADELTVRFAGLDURIDTKPDLTPVDADRAVESDVROTLGRDRPGDGVIGE 67
QY 140 ENGWRCACNSADPV---WVLDPIDGTSFKITGKPLFGTLIALHNGKPVIGVIDOPILRE 196
DB 68 EFG-----GSTFTTGQWIVDPIDGTFKFRVGPVWASLIALLEDGVPVGVVSAPALQR 122
QY 197 RWTVGDKGQ--TTLNG---QETSIVRSNCNLAQAYLYTTS-----PHLFEADAEDAFIRV 245
DB 123 RWAARGAFAGAFADGVRPHRLSVSSVAELHLSASLSFSSLSGWARPGL-----RERFIGL 177
QY 246 RNKV-KVPYLYGCDYAYALLAGFVDIVVESGLKPYDFLSVPVIEGAGGSITDWRGDKL 304
DB 178 TDTVWRVRYAG-DFLSYCLVAGAVDIAAEPQVSVWDLAALDITVREAGGRUITSLDG---- 233
QY 305 HWPVTAESRPTSFNVVAGDARVHKEALDAL 335

Db 234 -----VACPHGGSAVAT-NGLLHDEVLTRL 257

RESULT 13
F64103
suppressor protein subh - Haemophilus influenzae (strain Rd KW20)
C:Species: Haemophilus influenzae
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 28-Jul-2000
C:Accession: F64103
R: Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavac, J.; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidme, D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.; Science 269, 496-512, 1995
A: Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Vent, A.; Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A: Reference number: A64000; MUID: 95350630; PMID: 7542800
A: Accession: F64103
A: Status: nucleic acid sequence not shown; translation not shown
A: Molecule type: DNA
A: Residues: 1-267 <TIGR>
A: Cross-references: GB:U32775; GB:L42023; NID: g1573951; PIDN: AAC22595.1; PID: g1573951
C: Genetics:
C: Gene: subh
C: Superfamily: suppressor protein subh

Query Match 15.7%; Score 272; DB 1; Length 267;
Best Local Similarity 30.8%; Pred. No. 5.8e-14;
Matches 84; Conservative 44; Mismatches 113; Indels 32; Gaps 8;

QY 78 LVEVAQRAADAGEVLRKYFROR--VEIIDKEDHSPVTIADRAEAEAMVSVILKSPPTA 135
DB 4 MUNIAIRAARKAGNVIANKYERRDAIESTQKINDYVTVNDKASEAEIIEVIRKSYPDRT 63
QY 136 IFGEENGWRCACNSADFWVLDPIIDGTSFKITGKPLFGTLIALHNGKPVIGVIDOPILR 195
DB 64 ILTEETG-AIEGKDSQVQIIDLDTGRNFMGLPHFSVSIARVKNRTEVGVYDPIRN 122
QY 196 ERWIGVDGKQTLNGQEISVRSNCNLAQAYLYTTSPLH-----LFEADAEDAF-I 243
DB 123 ELFTVARGEGAKLNEVRI VDSKREIQSILATGFPFKPKLPKPAQAFAMNALIEDAADF 182
QY 244 RVNRKVVPYLYGCDYAYALLAGFVDIVVESGLKPYDFLSVPVIEGAGGSITDWRGDK 303
DB 183 RTGSAALDL----CY-----VASNRIDGVFENGKAWDCAGDLIVREAGGLVCFDF---- 231
QY 304 LHPVTAESRPTSFNVVAGDARVHKEALDALR 336
DB 232 -----AGNSYLSKSGNIIAA P'SKVIKEMLNKIR 258

RESULT 14
T07795
inositol-1(or 4)-monophosphatase (EC 3.1.3.25) isoform 3 - tomato
N: Alternate names: myo-inositol-1-phosphatase
C: Species: Lycopersicon esculentum (tomato)
C: Date: 14-May-1999 #sequence_revision 14-May-1999 #text_change 03-Jun-2002
C: Accession: T07795
R: Gillaspay, G.E.; Keddle, J.S.; Oda, K.; Gruijssem, W.
A: Title: Plant inositol monophosphatase is a lithium-sensitive enzyme encoded by a n Plant Cell 7, 2175-2185, 1995
A: Reference number: Z16138; MUID: 96351935; PMID: 8718627
A: Accession: T07795
A: Status: preliminary; translated from GB/EMBL/DBJ
A: Molecule type: mRNA
A: Residues: 1-268 <GII>
A: Cross-references: EMBL:U39059; NID: g1098970; PIDN: AAB19031.1; PID: g1098971
A: Experimental source: strain VENT Cherry
C: Genetics:
C: Gene: IMP3
C: Function:
A: Description: removes phosphate from myo-inositol-1-phosphate and myo-inositol 4-phosphate
A: Pathway: inositol phosphate metabolism

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 17, 2002, 16:39:23 ; Search time 20 seconds
(without alignments)
254.529 Million cell updates/sec

Title: US-09-686-522c-14
Perfect score: 1730
Sequence: 1 MLSSSTHSDTSFPGLAS.....VVAAGDARVHKEALDLRWR 338

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 97044 seqs, 15060890 residues

otal number of hits satisfying chosen parameters: 97044

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB pep.*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB pep.*
7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB pep.*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB pep.*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB pep.*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB pep.*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB pep.*

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	286	16.5	265	9	US-09-902-525-48
2	279	16.1	264	9	US-09-902-525-36
3	255.5	14.8	267	10	US-09-921-232-17
4	255.5	14.8	267	10	US-09-921-330-17
5	255.5	14.8	267	10	US-09-921-329-17
6	127	7.3	252	10	US-09-987-446-2
7	101	5.8	1616	9	US-09-712-363-262
8	93	5.4	268	10	US-09-216-393-9C
9	90	5.2	667	10	US-09-896-832-55
10	90	5.2	777	10	US-09-925-301-1115
11	88	5.1	3472	9	US-10-027-806-4
12	86	5.0	912	10	US-09-817-464-2
13	84	4.9	3739	10	US-09-860-846-33
14	84	4.9	3739	10	US-09-861-289-33
15	84	4.9	11877	9	US-09-860-846-6
16	84	4.9	11877	10	US-09-861-289-6
17	83.5	4.8	340	10	US-09-159-469-54
18	83.5	4.8	340	10	US-09-798-042-54
19	83.5	4.8	2383	10	US-09-912-020-302

20	83	4.8	312	10	US-09-815-242-5213	Sequence 5213, Ap
21	83	4.8	499	10	US-09-815-242-11025	Sequence 11025, A
22	82.5	4.8	598	10	US-09-815-242-13230	Sequence 13230, A
23	82.5	4.8	792	10	US-09-815-242-11966	Sequence 11966, A
24	82.5	4.8	943	10	US-09-815-242-11834	Sequence 11834, A
25	82.5	4.8	4613	9	US-09-860-846-31	Sequence 31, Appl
26	82.5	4.8	4613	10	US-09-861-289-31	Sequence 31, Appl
27	82	4.7	799	10	US-09-912-917-6	Sequence 6, Appl1
28	82	4.7	953	10	US-09-912-917-2	Sequence 2, Appl1
29	81.5	4.7	943	10	US-09-815-242-12027	Sequence 12027, A
30	80.5	4.7	710	10	US-09-801-368-276	Sequence 276, App
31	80	4.6	663	10	US-09-815-242-11869	Sequence 11869, A
32	79.5	4.6	509	9	US-10-121-032-18	Sequence 18, Appl
33	79	4.6	766	10	US-09-925-301-1276	Sequence 1276, Ap
34	79	4.6	820	10	US-09-815-242-10771	Sequence 10771, A
35	78.5	4.5	409	10	US-09-815-242-11270	Sequence 11270, A
36	78.5	4.5	639	10	US-09-854-731-17	Sequence 17, Appl
37	78	4.5	437	10	US-09-815-242-11187	Sequence 11187, A
38	78	4.5	681	10	US-09-815-242-11830	Sequence 11830, A
39	78	4.5	2785	10	US-09-801-574-8	Sequence 8, Appl1
40	77	4.5	496	10	US-09-945-301-4	Sequence 4, Appl1
41	76.5	4.4	1114	12	US-10-005-983-2	Sequence 2, Appl1
42	76	4.4	329	10	US-09-925-300-1406	Sequence 1406, Ap
43	76	4.4	446	12	US-10-081-859-2	Sequence 2, Appl1
44	75.5	4.4	280	9	US-09-960-631A-5	Sequence 5, Appl1
45	75.5	4.4	294	10	US-09-801-368-202	Sequence 202, App

ALIGNMENTS

RESULT 1
US-09-902-525-48
; Sequence 48, Application US/09902525
; Patent No. US20020164751A1
; GENERAL INFORMATION:
; APPLICANT: Recombinant Biocatalysis, Inc.
; TITLE OF INVENTION: THERMOSTABLE PHOSPHATAS;
; NUMBER OF SEQUENCES 54
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows 95
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/902,525
; FILING DATE: 09-Jul-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/202,681A
; FILING DATE: 23-Dec-1999
; APPLICATION NUMBER: 08/666,857
; FILING DATE: 19-JUN-1996
; APPLICATION NUMBER: 60/033,752
; FILING DATE: 19-JUN-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Halle, Ph.D., Lisa A.
; REGISTRATION NUMBER: 38,347
; REFERENCE/DOCKET NUMBER: 09010/015W01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619/678-5070
; TELEFAX: 619/678-5099
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 48:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 265 amino acids

Db 61 DHKFIGEESAALGATADLTDDPTWIVDPLDGTTFNVHGFPPFCVSVGLTGKTIPTVGVW 120
Qy 190 DOPILRERWIGVDGKQTTLNGQEISVRSNCNLLAQAYLYTTSPHLFEADAEDAFIRVNKV 249
Db 121 FNPIMNELFTAVRGKGAFLNGSPKASSQDELVKALLVTEAGTNRDKTTVD---DTTNRI 177
Qy 250 KVPYLGCD---CYAYAL----LASGFVDIVVESGL-KPYDFLSLVPVIEGAGGSITDWR 300
Db 178 NRLLYKIRSIKRMCSLALNMGCVACGRDLDCYEGFGGPDVDAAGAVILQEGAGGLVFDPS 237
Qy 301 G 301
Db 238 G 238

RESULT 4
US-09-921-330-17
; Sequence 17, Application US/09921330
; Patent No. US20020102682A1
GENERAL INFORMATION:
; APPLICANT: Martino-Catt, Susan J.
; APPLICANT: Wang, Hongyu
; TITLE OF INVENTION: Polypeptides Controlling Phytate
; TITLE OF INVENTION: Metabolism in Plants
; FILE REFERENCE: 0706D3
; CURRENT FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: 60/055,446
; PRIOR FILING DATE: 1997-08-11
; PRIOR APPLICATION NUMBER: 60/055,526
; PRIOR FILING DATE: 1997-08-08
; PRIOR APPLICATION NUMBER: 60/053,944
; PRIOR FILING DATE: 1997-07-28
; PRIOR APPLICATION NUMBER: 09/118,442
; PRIOR FILING DATE: 1998-07-17
; PRIOR APPLICATION NUMBER: 09/677,064
; PRIOR FILING DATE: 2000-09-29
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 17
; LENGTH: 267
; TYPE: PRT
; ORGANISM: Zea mays
US-09-921-330-17

Query Match 14.8%; Score 255.5; DB 10; Length 267;
Best Local Similarity 30.3%; Pred. No. le-16;
Matches 73; Conservative 41; Mismatches 112; Indels 15; Gaps 5;
Qy 73 LATERLVEAQRADAAGEVLRYKRYFRORVEIIDKEDHSPVTIADREAEAMVSVILKSFP 132
Db 1 MSEEQFLAVAVEAAKSAGEIIRKGFYQTKNVQHGQVDLVTETDKACEDLIFNHLRKHFP 60
Qy 133 THAIFGEENG---WRCAENSADFVWVLDPIDGTSFKITGKPLFGTLTALLHNGKPVIGVI 189
Db 61 DHKFIGEESAALGATADLTDDPTWIVDPLDGTTFNVHGFPPFCVSVGLTGKTIPTVGVW 120
Qy 190 DOPILRERWIGVDGKQTTLNGQEISVRSNCNLLAQAYLYTTSPHLFEADAEDAFIRVNKV 249
Db 121 FNPIMNELFTAVRGKGAFLNGSPKASSQDELVKALLVTEAGTNRDKTTVD---DTTNRI 177
Qy 250 KVPYLGCD---CYAYAL----LASGFVDIVVESGL-KPYDFLSLVPVIEGAGGSITDWR 300
Db 178 NRLLYKIRSIKRMCSLALNMGCVACGRDLDCYEGFGGPDVDAAGAVILQEGAGGLVFDPS 237
Qy 301 G 301
Db 238 G 238
RESULT 5

US-09-921-329-17
; Sequence 17, Application US/09921329
; Patent No. US20020110884A1
GENERAL INFORMATION:
; APPLICANT: Martino-Catt, Susan J.
; APPLICANT: Wang, Hongyu
; APPLICANT: Beach, Larry R.
; TITLE OF INVENTION: Polypeptides Controlling Phytate
; TITLE OF INVENTION: Metabolism in Plants
; FILE REFERENCE: 0706D2
; CURRENT FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: 60/055,446
; PRIOR FILING DATE: 1997-08-11
; PRIOR APPLICATION NUMBER: 60/055,526
; PRIOR FILING DATE: 1997-08-08
; PRIOR APPLICATION NUMBER: 60/053,944
; PRIOR FILING DATE: 1997-07-28
; PRIOR APPLICATION NUMBER: 09/118,442
; PRIOR FILING DATE: 1998-07-17
; PRIOR APPLICATION NUMBER: 09/677,064
; PRIOR FILING DATE: 2000-09-29
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 17
; LENGTH: 267
; TYPE: PRT
; ORGANISM: Zea mays
US-09-921-329-17

Query Match 14.8%; Score 255.5; DB 10; Length 267;
Best Local Similarity 30.3%; Pred. No. le-16;
Matches 73; Conservative 41; Mismatches 112; Indels 15; Gaps 5;
Qy 73 LATERLVEAQRADAAGEVLRYKRYFRORVEIIDKEDHSPVTIADREAEAMVSVILKSFP 132
Db 1 MSEEQFLAVAVEAAKSAGEIIRKGFYQTKNVQHGQVDLVTETDKACEDLIFNHLRKHFP 60
Qy 133 THAIFGEENG---WRCAENSADFVWVLDPIDGTSFKITGKPLFGTLTALLHNGKPVIGVI 189
Db 61 DHKFIGEESAALGATADLTDDPTWIVDPLDGTTFNVHGFPPFCVSVGLTGKTIPTVGVW 120
Qy 190 DOPILRERWIGVDGKQTTLNGQEISVRSNCNLLAQAYLYTTSPHLFEADAEDAFIRVNKV 249
Db 121 FNPIMNELFTAVRGKGAFLNGSPKASSQDELVKALLVTEAGTNRDKTTVD---DTTNRI 177
Qy 250 KVPYLGCD---CYAYAL----LASGFVDIVVESGL-KPYDFLSLVPVIEGAGGSITDWR 300
Db 178 NRLLYKIRSIKRMCSLALNMGCVACGRDLDCYEGFGGPDVDAAGAVILQEGAGGLVFDPS 237
Qy 301 G 301
Db 238 G 238

RESULT 6
US-09-987-446-2
; Sequence 2, Application US/09987446
; Patent No. US20020115162A1
GENERAL INFORMATION:
; APPLICANT: FARWICK, Mike, et al.
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCES CODING FOR THE cysQ GENE
; FILE REFERENCE: 032301 WD 245
; CURRENT APPLICATION NUMBER: US/09/987,446
; CURRENT FILING DATE: 2001-11-14
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 252
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-987-446-2

```
Query Match          7.3%; Score 127; DB 10; Length 252;
Best Local Similarity 28.2%; Pred. No. 0.00013;
Matches 37; Conservative 21; Mismatches 61; Indels 12; Gaps 4;

Qy 81 VAQRAADAAGEVLKRYFRQVEIIDKEDHSPVTIADREAEAMVSVILKSFPTTHAIFGEE 140
Db 10 LTHRLAAGTGEILKGV--RNWGLVGRNLGDA--GDELAQSWIARVLEQHRPNDGFLSEE 65

Qy 141 NGWRCAN-----SADFVWLDPIDGTSFITGPKPLGFTGLIALHNGKPVIGVIDQPIRE 196
Db 66 ----AADNDRLSKDRWIIIDPLDGTKEFATGRQDMVAVHIALVENGVPVTHAAVGLPDLGV 121

Qy 197 RWIGVGCKQTT 207
Db 122 VFHSADARAVT 132

RESULT 7
US-09-712-363-262
Sequence 262, Application US/09712363
Patent No. US20020164588A1
GENERAL INFORMATION:
; APPLICANT: Eisenberg, David
; APPLICANT: Rotstein, Sergio H.
; APPLICANT: Marcotte, Edward M.
; TITLE OF INVENTION: DETERMINING THE FUNCTIONS AND
; TITLE OF INVENTION: INTERACTIONS OF PROTEINS BY COMPARATIVE ANALYSIS
; FILE REFERENCE: 07419-032001
; CURRENT APPLICATION NUMBER: US/09/712,363
; PRIOR FILING DATE: 2000-11-13
; PRIOR APPLICATION NUMBER: PCT/US00/02246
; PRIOR FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/179,531
; PRIOR FILING DATE: 2000-02-01
; PRIOR APPLICATION NUMBER: 60/117,844
; PRIOR FILING DATE: 1999-01-29
; PRIOR APPLICATION NUMBER: 60/118,206,
; PRIOR FILING DATE: 1999-02-01
; PRIOR APPLICATION NUMBER: 60/126,593
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 60/134,093
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 60/134,092
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 60/165,124
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/165,086
; PRIOR FILING DATE: 1999-11-12
NUMBER OF SEQ ID NOS: 292
SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 262
; LENGTH: 1616
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-712-363-262

Query Match          5.8%; Score 101; DB 9; Length 1616;
Best Local Similarity 24.2%; Pred. No. 0.56;
Matches 76; Conservative 32; Mismatches 116; Indels 90; Gaps 17;

Qy 38 VSSAVLSASG-----RQPMSTVRASFAAGAAGR---RAAAGVGLATERLVEVA 82
Db 638 VLDVAVLHALGLAVLKEKTAQSTETRLPFCWGRVSLHAGGAGRVARFASAGADA-----IS 691

Qy 83 QRAADAAG-EVLKRYFRQVEIIDKEDHSPVTIADREAEAMVSVILKSFPTTHAIFGEE 141
Db 692 VDVCATGLPVLTVRSLSVTRFITAQLRAAVTAAGGASDQGLEVVWS--PISVVSGGAN 749

Qy 142 G-----WR--CAENSAD----FVWVLDPIDGTSKITGKPLGFTGLIALHNGKPVIGV 188
Db 750 GSAPPAPVSWADFCAAGSDGASVVVWVELESAGGQASVV-----GSVYAATHTALEVL-- 802

Qy 189 IDQPILRERWIGVGKQTT-----LNQSEIVSRSCNLLAQAYLYTTTSPHLFEADAE 239
```

```
Db 803 -----QSWLGADRAATLVILTHGCVGLAGEDISD-----LAAAVWGMAR---SAQAE 847
Qy 240 DAFIRVRNKVKVPLXGDCDY--AYALLASGFDVIIVESG-----LKPYDFLSLVPVIE- 290
Db 848 -----NPGRIVLIDTDAADVADSVLAGVGEPLLVRGGTVHAPRLSPAPALLALPAAES 900
Qy 291 -----GAGGSITD 298
Db 901 AWRLAAGGGGTLED 914

RESULT 8
US-09-216-393-90
; Sequence 90, Application US/09216393
; Patent No. US20010014447A1
; GENERAL INFORMATION:
; APPLICANT: Milhausen, Michael James
; TITLE OF INVENTION: TOXOPLASMA GONDII PROTEINS, NUCLEIC ACID MOLECULES, AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: TX-1-C2
; CURRENT APPLICATION NUMBER: US/09/216,393
; CURRENT FILING DATE: 1998-12-18
; EARLIER APPLICATION NUMBER: 08/994,825
; EARLIER FILING DATE: 1997-12-19
; NUMBER OF SEQ ID NOS: 364
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 90
; LENGTH: 268
; TYPE: PRT
; ORGANISM: Toxoplasma gondii
US-09-216-393-90

Query Match          5.4%; Score 93; DB 10; Length 268;
Best Local Similarity 22.7%; Pred. No. 0.24;
Matches 42; Conservative 32; Mismatches 53; Indels 59; Gaps 8;

Qy 16 PGLASANPNRSLRLRLRAASPVSSAVLSAS-----GRQPMSTVRASFAAGAAGR 65
Db 7 PTLISMN---QSTIVATSVVAQSAVLSRAPSRPGPSESFGKQESRPGVSGAGLAESK 63

Qy 66 RAAAVGELATERLVEVAQRAADAAGEVLKRYFRQVEIIDKEDHSPVTIADREAEAMVS 125
Db 64 RVPSLTQPSLERSVTISRRIKIDAVG-----MSLVFKLDRTTTSLAAKEEFSSID 113

Qy 126 VILKSFPTTHAIFGEE-----GW-----RCAENSADFVWVLDPIDG 161
Db 114 KIV-SKPTHS-FCESKLPAGIMKAKSMFPQSQTLSAPWNAPARCARKDS-----FG 162

Qy 162 TKSFI 166
Db 163 TKAWI 167

RESULT 9
US-09-896-852-55
; Sequence 55, Application US/09896852
; Patent No. US200020025542A1
; GENERAL INFORMATION:
; APPLICANT: Abbott Laboratories
; APPLICANT: Maine, Gregory T.
; APPLICANT: Hunt, Jeffery C.
; APPLICANT: Brojanac, Susan
; APPLICANT: Jhy-Tsing Sheu, Michael
; APPLICANT: Chovan, Linda E.
; APPLICANT: Tyner, Joan D.
; APPLICANT: Howard, Lawrence V.
; TITLE OF INVENTION: ANTIGEN COCKTAILS, P35, AND USES THEREOF
; FILE REFERENCE: 6361 US.D1
; CURRENT APPLICATION NUMBER: US/09/896,852
; CURRENT FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 09/086,503
; PRIOR FILING DATE: 1998-05-28
```

RESULT 12
US-09-817
; Sequenc

Query Match 4.9%; Score 84; DB 9; Length 11877;
Best Local Similarity 32.3%; Pred. No. 4e+02;
Matches 31; Conservative 15; Mismatches 40; Indels 10; Gaps 4;
QY 46 SGRQPMSTVRASFAAGAAGRRAA--AVGELATERLVEVAQRAADA--AGEVLKRYFRQRV 101
DB 8220 SGLDPEGTVLLTGGTGALGGIVARHVVGWGVRRLLLVSRRTDAPGAGELYHE-----L 8274
QY 102 EIIDKE-DHSPVTIADREAEAMVSVILKSFPTHAI 136
DB 8275 EALGADVSAACDVADREALTAVLDSIPAETHPLTAV 8310

Search completed: November 17, 2002, 16:44:49
Job time : 56 secs